PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/82, 9/10, 1/00, 15/86, G01N 33/50, C12Q 1/68, A01H 5/00 — A2

(11) International Publication Number: WO 00/24915

(43) International Publication Date: 4 May 2000 (04.05.00)

(21) International Application Number: PCT/US99/24923

(22) International Filing Date: 22 October 1999 (22.10.99)

(30) Priority Data:
60/105,451 23 October 1998 (23.10.98) US

(71) Applicant *(for all designated States except US)*: E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: ALLEN, Stephen, M. [US/US]; 2225 Rosewood Drive, Wilmington, DE 19810 (US). HITZ, William, D. [US/US]; 404 Hillside Road, Wilmington, DE 19807 (US).

(74) Agent: FEULNER, Gregory, J.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AE, AL, AU, BA, BB, BG, BR, CA, CN, CR, CU, CZ, DM, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PLANT RAFFINOSE SYNTHASE HOMOLOGS

(57) **Abstract**

This invention relates to an isolated nucleic acid fragment encoding a raffinose synthase. This invention also relates to the construction of a chimeric gene encoding all or a portion of the raffinose synthase, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the raffinose synthase in a transformed host cell.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TITLE

PLANT RAFFINOSE SYNTHASE HOMOLOGS

This application claims the benefit of U.S. Provisional Application No. 60/105,451, filed October 23, 1998.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding raffinose synthase homologs in plants and seeds.

BACKGROUND OF THE INVENTION

Raffinose saccharides are a group of D-galactose-containing oligosaccharides of sucrose that are widely distributed in plants. Raffinose saccharides are characterized by having the general formula: [0-β-D-galactopyranosyl-$(1\rightarrow6)_n$-α-glucopyranosyl-(1→2)-β-D-fructofuranoside where n=0 through n=4 are known respectively as sucrose, raffinose, stachyose, verbascose, and ajugose.

Extensive botanical surveys of the occurrence of raffinose saccharides have been reported in the scientific literature (see Dey (1985) in Biochemistry of Storage Carbohydrates in Green Plants, Academic Press, London, pp. 53-129). Raffinose saccharides are thought to be second only to sucrose among the nonstructural carbohydrates with respect to abundance in the plant kingdom. In fact, raffinose saccharides may be ubiquitous, at least among higher plants. Raffinose saccharides accumulate in significant quantities in the edible portion of many economically significant crop species. Examples include soybean (*Glycine max* L. Merrill), sugar beet (*Beta vulgaris*), cotton (*Gossypium hirsutum* L.), canola (*Brassica* sp.) and all of the major edible leguminous crops including beans (*Phaseolus* sp.), chick pea (*Cicer arietinum*), cowpea (*Vigna unguiculata*), mung bean (*Vigna radiata*), peas (*Pisum sativum*), lentil (*Lens culinaris*) and lupine (*Lupinus* sp.).

The biosynthesis of raffinose saccharides has been fairly well characterized (see Dey (1985) in Biochemistry of Storage Carbohydrates in Green Plants). The committed reaction of raffinose saccharide biosynthesis involves the synthesis of galactinol (O-α-D-galactopyranosyl-(1→1)-myo-inositol) from UDPgalactose and myo-inositol. The enzyme that catalyzes this reaction is galactinol synthase. Synthesis of raffinose and higher homologues in the raffinose saccharide family from sucrose is thought to be catalyzed by distinct galactosyltransferases (e.g., raffinose synthase, stachyose synthase, etc.).

Although abundant in many species, raffinose saccharides are an obstacle to the efficient utilization of some economically important crop species. Raffinose saccharides are not digested directly by animals, primarily because α-galactosidase is not present in the intestinal mucosa (Gitzelmann and Auricchio (1965) *Pediatrics 36*:231-236; Rutloff et al. (1967) *Nahrung 11*:39-46). However, microflora in the lower gut are readily able to

ferment the raffinose saccharides which results in an acidification of the gut and production of carbon dioxide, methane and hydrogen (Murphy et al. (1972) *J. Agr. Food Chem. 20*:813-817; Cristofaro et al. (1974) in Sugars in Nutrition, Ch 20, 313-335; Reddy et al. (1980) *J. Food Science 45*:1161-1164). The resulting flatulence can severely limit the use of leguminous plants in animal, including human, diets. It is unfortunate that the presence of raffinose saccharides restricts the use of soybeans in animal, including human, diets because otherwise this species is an excellent source of protein and fiber.

The problems and costs associated with raffinose saccharides could be reduced or eliminated through the availability of genes that confer a reduction of raffinose saccharide content of soybean seeds. Such genes could be used to develop soybean varieties having inherently reduced raffinose saccharide content. Soybean varieties with inherently reduced raffinose saccharide content would improve the nutritional quality of derived soy protein products and reduce processing costs associated with the removal of raffinose saccharides. Said low raffinose saccharide soybean varieties would be more valuable than conventional varieties for animal and human diets and would allow mankind to more fully utilize the desirable nutritional qualities of this edible legume.

Imbibition proteins from barley, *Brassica oleracea*, *Arabidopsis thaliana*, and *Cicer arietinum* are a group of uncharacterized proteins found in swelling seeds. Interestingly, raffinose synthase is expressed very late in seed maturation and the protein or its mRNA may be present in seeds after drydown. Thus raffinose synthase may be a imbibition protein.

In light of the above described factors, it is apparent that soybean plants with heritable, substantially reduced raffinose saccharide content useful for preparing soy protein products with an improved carbohydrtae content are needed. Heretofore, the only means to achieve a desirable raffinose saccharide content was to physically and/or chemically treat the soybean. Thus, there is a great deal of interest in identifying the genes that encode proteins involved in raffinose saccharide biosynthesis in plants. These genes may be used in plant cells to alter levels of raffinose biosynthesis. Accordingly, the availability of nucleic acid sequences encoding all or a portion of raffinose saccharide biosynthetic proteins would facilitate studies to better understand cellular metabolism and raffinose production in plants, provide genetic tools to manipulate cellular metabolism and alter raffinose production.

SUMMARY OF THE INVENTION

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a first polypeptide of at least 750 amino acids that has at least 85% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of corn polypeptides of SEQ ID NOs:2 and 4, rice polypeptides of SEQ ID NOs:6 and 8, soybean polypeptide of SEQ ID NO:10 and a wheat polypeptide of

SEQ ID NO:16. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

The present invention further relates to an isolated polynucleotide comprising a nucleotide sequence encoding a second polypeptide of at least 45 amino acids that has at least 70% identity based on the Clustal method of alignment when compared to a soybean polypeptide of SEQ ID NO:12. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

The present invention further relates to an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide of at least 750 amino acids that has at least 95% identity based on the Clustal method of alignment when compared to a wheat polypeptide of SEQ ID NO:14. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

It is preferred that the isolated polynucleotides of the claimed invention consists of nucleic acid sequences selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13 and 15 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14 and 16. The present invention also relates to an isolated polynucleotide comprising a nucleotide sequences of at least one of 40 (preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13 and 15 and the complement of such nucleotide sequences.

The present invention relates to a chimeric gene comprising an isolated polynucleotide of the present invention operably linked to suitable regulatory sequences.

The present invention relates to an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention. The host cell may be eukaryotic, such as a yeast or a plant cell, or prokaryotic, such as a bacterial cell or a virus. If the host cell is a virus, it is preferably a baculovirus. A baculovirus comprising an isolated polynucleotide of the present invention or a chimeric gene of the present invention is most preferred.

The present invention relates to a process for producing an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention, the process comprising either transforming or transfecting an isolated compatible host cell with a chimeric gene or isolated polynucleotide of the present invention.

The present invention relates to a raffinose synthase polypeptide of at least 750 amino acids comprising at least 85% homology based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, and 16.

The present invention relates to a raffinose synthase polypeptide having at least 70% identity based on the Clustal method of alignment when compared to the polypeptide of SEQ ID NO:12.

The present invention also relates to a raffinose synthase polypeptide of at least 750 amino acids comprising at least 95% homology based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:14.

The present invention relates to a method of selecting an isolated polynucleotide that affects the level of expression of a raffinose synthase polypeptide in a plant cell, the method comprising the steps of:

constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention;

introducing the isolated polynucleotide or the isolated chimeric gene into a plant cell;

measuring the level a raffinose synthase polypeptide in the plant cell containing the isolated polynucleotide; and

comparing the level of a raffinose synthase polypeptide in the plant cell containing the isolated polynucleotide with the level of a raffinose synthase polypeptide in a plant cell that does not contain the isolated polynucleotide.

The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a raffinose synthase polypeptide gene, preferably a plant raffinose synthase polypeptide gene, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 30 (preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13 and 15 and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of a raffinose synthase amino acid sequence.

The present invention also relates to a method of obtaining a nucleic acid fragment encoding all or a subsantial portion of the amino acid sequence encoding a raffinose synthase polypeptide comprising the steps of: probing a cDNA or genomic library with an isolated polynucleotide of the present invention; identifying a DNA clone that hybridizes with an isolated polynucleotide of the present invention; isolating the identified DNA clone; and sequencing the cDNA or genomic fragment that comprises the isolated DNA clone.

The present invention also relates to an expression cassette comprising an isolated polynucleotide of the present invention operably linked to a promoter or a chimeric gene of the present invention.

The present invention also relates to a method of positive selection of a transformed cell comprising transforming a plant cell with the chimeric gene or the expression cassette of

the present invention; and growing the transformed plant under conditions allowing expression of the polynucleotide (such as raffinose synthase) in an amount sufficient to reduce raffinose saccharide content of the plant to provide a positive selection means. Preferably the plant cell is a dicot.

BRIEF DESCRIPTION OF THE SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying Sequence Listing which form a part of this application.

Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

TABLE 1

Raffinose Synthase Homologs

Protein	Clone Designation	SEQ ID NO: (Nucleotide)	SEQ ID NO: (Amino Acid)
Raffinose Synthase	cbn10.pk0054.d4	1	2
Raffinose Synthase	Contig composed of cbn10.pk0034.e8 chpc24.pk0003.h7	3	4
Raffinose Synthase	rls24.pk0017.g10	5	6
Raffinose Synthase	rls72.pk0020.d9	7	8
Raffinose Synthase	sfl1.pk125.d4	9	10
Raffinose Synthase	sgs2c.pk005.c14	11	12
Raffinose Synthase	wlm24.pk0021.h1	13	14
Raffinose Synthase	wlm96.pk033.h5	15	16

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Res. 13*:3021-3030 (1985) and in the *Biochemical J. 219 (No. 2)*:345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, a "polynucleotide" is a nucleotide sequence such as a nucleic acid fragment. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally

contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, or synthetic DNA. An isolated polynucleotide of the present invention may include at least one of 40 contiguous nucleotides, preferably at least one of 30 contiguous nucleotides, most preferably one of at least 15 contiguous nucleotides, of the nucleic acid sequence of the SEQ ID NOs:1, 3, 5, 7, 9, 11, 13 and 15.

As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by gene silencing through for example antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof.

Substantially similar nucleic acid fragments may be selected by screening nucleic acid fragments representing subfragments or modifications of the nucleic acid fragments of the instant invention, wherein one or more nucleotides are substituted, deleted and/or inserted, for their ability to affect the level of the polypeptide encoded by the unmodified nucleic acid fragment in a plant or plant cell. For example, a substantially similar nucleic acid fragment representing at least one of 30 contiguous nucleotides derived from the instant nucleic acid fragment can be constructed and introduced into a plant or plant cell. The level of the polypeptide encoded by the unmodified nucleic acid fragment present in a plant or plant cell exposed to the substantially similar nucleic fragment can then be compared to the level of the polypeptide in a plant or plant cell that is not exposed to the substantially similar nucleic acid fragment. For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments

representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Consequently, an isolated polynucleotide comprising a nucleotide sequence of at least one of 40 (preferably at least one of 30, most preferably at least one of 15) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13 and 15 and the complement of such nucleotide sequences may be used in methods of selecting an isolated polynucleotide that affects the expression of a polypeptide (such as raffinose synthase) in a host cell, preferably a plant cell. A method of selecting an isolated polynucleotide that affects the level of expression of a polypeptide in a host cell (eukaryotic, such as plant, or prokarotic such as yeast bacterial or virus) may comprise the steps of: constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention; introducing the isolated polynucleotide or the isolated chimeric gene into a host cell; measuring the level a polypeptide in the host cell containing the isolated polynucleotide; and comparing the level of a polypeptide in the host cell containing the isolated polynucleotide with the level of a polypeptide in a host cell that does not contain the isolated polynucleotide.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) Nucleic Acid Hybridisation, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with

0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are 95% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above homologies but typically encode a polypeptide having at least 50 amino acids, preferably 100 amino acids, more preferably 150 amino acids, still more preferably 200 amino acids, and most preferably 250 amino acids. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or

more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that

are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a nucleotide sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants 15*:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Mol. Biotechnol. 3*:225-236).

The "3' non-coding sequences" refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell 1*:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into polypeptide by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to an RNA transcript that includes the mRNA and so can be translated into a polypeptide by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys. 100*:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol. 143*:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London) 327*:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several raffinose synthase homologs have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other raffinose synthase, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA 85*:8998-9002) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl. Acad. Sci. USA 86*:5673-5677; Loh et al. (1989) *Science 243*:217-220). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques 1*:165). Consequently, a polynucleotide comprising a nucleotide sequence of at least one of 40 (preferably one of at least 30, most preferably one of at least 15) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15 and the complement of such nucleotide sequences may be used in such methods to obtain a nucleic acid fragment encoding a substantial portion of an amino acid sequence of a polypeptide (such as raffinose synthase). The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a polypeptide of a gene (such as raffinose synthase) preferably a substantial portion of a plant polypeptide of a gene, comprising the steps of : synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of

40 (preferably at least one of 30, most preferably at least one of 15) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13 and 15 and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of a polypeptide (such as raffinose synthase).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol. 36*:1-34; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of raffinose synthase in those cells.

Overexpression of the proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J. 4*:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics 218*:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptides to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by altering the coding

sequence to encode the instant polypeptides with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell 56*:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.100*:1627-1632) added and/or with targeting sequences that are already present removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the instant polypeptides in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant polypeptide can be constructed by linking a gene or gene fragment encoding that polypeptide to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U.S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In addition, the ability to restrict the expression of specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic advantages relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppression technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on practical grounds, and is not an inherent part of the invention. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein

encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be one which allows large numbers of samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

The instant polypeptides (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptides. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded raffinose synthase homolog. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 6).

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics 1*:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet. 32*:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter 4*:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; *see* Hoheisel et al. In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) *Trends Genet. 7*:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan et al. (1995) *Genome Res. 5*:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med. 11*:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics 16*:325-332), allele-specific ligation (Landegren et al. (1988) *Science 241*:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res. 18*:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nat. Genet. 7*:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res. 17*:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer (1989) *Proc. Natl. Acad. Sci USA 86*:9402-9406; Koes et al. (1995) *Proc. Natl. Acad. Sci USA 92*:8149-8153; Bensen et al. (1995) *Plant Cell 7*:75-84). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the instant polypeptides. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With

either method, a plant containing a mutation in the endogenous gene encoding the instant polypeptides can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the instant polypeptides disclosed herein.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean and Wheat

Library	Tissue	Clone
cbn10	Corn developing kernel (embryo and endosperm); 10 days after pollination	cbn10.pk0054.d4 cbn10.pk0034.e8
chpc24	Corn 8 day old shoot treated 24 hours with herbicide*	chpc24.pk0003.h7
rls24	Rice leaf 15 days after germination, 24 hours after infection of strain *Magaporthe grisea* 4360-R-67 (AVR2-YAMO)	rls24.pk0017.g10
rls72	Rice leaf 15 days after germination, 72 hours after infection of strain *Magaporthe grisea* 4360-R-67 (AVR2-YAMO)	rls72.pk0020.d9
sfl1	Soybean immature flower	sfl1.pk125.d4
sgs2c	Soybean seeds 14 hours after germination	sgs2c.pk005.c14
wlm24	Wheat seedlings 24 hours after inoculation with *Erysiphe graminis f. sp tritici*	wlm24.pk0021.h1
wlm96	Wheat seedlings 96 hours after inoculation with *Erysiphe graminis f. sp tritici*	wlm96.pk033.h5

*Application of 2-[(2,4-dihydro-2,6,9-trimethyl[1]benzothiopyrano[4,3-*c*]pyrazol-8-yl)carbonyl]-1,3-cyclohexanedione *S,S*-dioxide; synthesis and methods of using this compound are described in WO 97/19087, incorporated herein by reference

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene

Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science 252*:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

cDNA clones encoding raffinose synthase homologs were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet. 3*:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding Raffinose Synthase Homologs

The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to raffinose synthase from *Arabidopsis thaliana* (NCBI Identifier No. gi 3959466), *Hordeum vulgare* (NCBI Identifier No. gi 282994), *Cucumis sativus* (NCBI Identifer No. 4106395) and *Brassica oleracea*

(NCBI Identifier No. gi 629602). The BLASTX search using the EST sequence from clone sgs2c.pk005.c14 revealed similarity of the protein encoded by the cDNA to a probable imbibition protein from *Hordeum vulgare* (NCBI Identifier No. gi 167100). Imbibition proteins from *Hordeum vulgare*, *Brassica oleracea*, *Arabidopsis thaliana*, and *Cicer arietinum* are a group of uncharacterized proteins found in swelling seeds. Raffinose synthase is expressed very late in seed maturation and the protein or its mRNA may be present in seeds after drydown. Thus raffinose synthase may also be an imbibition protein.

Shown in Table 3 are the BLAST results for individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), contigs assembled from two or more ESTs ("Contig"), contigs assembled from an FIS and one or more ESTs ("Contig*"), or sequences encoding the entire protein derived from an FIS, a contig, or an FIS and PCR ("CGS"):

TABLE 3

BLAST Results for Sequences Encoding Polypeptides Homologous to *Arabidopsis thaliana*, *Hordeum vulgare*, *Cucumis sativus* and *Brassica oleracea* Raffinose Synthase

Clone	Status	BLAST pLog Score
cbn10.pk0054.d4	FIS	>254.00 (gi 3953466)
Contig composed of cbn10.pk0034.e8 chpc24.pk0003.h7	Contig	>254.00 (gi 282994)
rls24.pk0017.g10	FIS	>254.00 (gi 282994)
rls72.pk0020.d9	FIS	>254.00 (gi 282994)
sfl1.pk125.d4	FIS	>254.00 (gi 4106395)
sgs2c.pk005.c14	EST	7.70 (gi 167100)
wlm24.pk0021.h1	FIS	>254.00 (gi 282994)
wlm96.pk033.h5	FIS	>254.00 (gi 282994)

The data in Table 4 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14 and 16 and the *Arabidopsis thaliana*, *Hordeum vulgare*, *Cucumis sativus* and *Brassica oleracea* sequences. The percent identity between each of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14 and 16 ranged from 14% to 82%.

TABLE 4

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to *Arabidopsis thaliana*, *Hordeum vulgare*, *Cucumis sativus* and *Brassica oleracea* Raffinose Synthase

SEQ ID NO.	Percent Identity to
2	63% (gi 3953466)
4	80% (gi 282994)
6	60% (gi 282994)
8	81% (gi 282994)
10	67% (gi 4106395)
12	57% (gi 167100)
14	94% (gi 282994)
16	54% (gi 282994)

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of a raffinose synthase. These sequences represent the first corn, rice, and wheat sequences and a new soybean sequence encoding raffinose synthase.

EXAMPLE 4

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas,

VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 μm in diameter) are coated with DNA using the following technique. Ten μg of plasmid DNAs are added to 50 μL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 μL of a 2.5 M solution) and spermidine free base (20 μL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles

resuspended in a final volume of 30 μL of ethanol. An aliquot (5 μL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology 8*:833-839).

EXAMPLE 5

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem. 261*:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described

above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embroys may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) *327*:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature 313*:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al.(1983) *Gene 25*:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL $CaCl_2$ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the

retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 6

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene 56*:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 μg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 μL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 μg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides

are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol. 189*:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

CLAIMS

What is claimed is:

1. A composition comprising an isolated polynucleotide comprising a nucleotide sequence encoding a first polypeptide of at least 750 amino acids that has at least 85% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a polypeptide of SEQ ID NOs:2, 4, 6, 8, 10 and 16 or an isolated polynucleotide comprising the complement of the nucleotide sequence.

2. The composition of Claim 1, wherein the isolated nucleotide sequence consists of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9 and 15 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10 and 16.

3. The composition of Claim 1 wherein the isolated polynucleotide is DNA.

4. The composition of Claim 1 wherein the isolated polynucleotide is RNA.

5. A chimeric gene comprising the isolated polynucleotide of Claim 1 operably linked to suitable regulatory sequences.

6. An isolated host cell comprising the chimeric gene of Claim 5.

7. An isolated host cell comprising an isolated polynucleotide of Claim 1 or Claim 3.

8. The isolated host cell of Claim 7 wherein the isolated host is selected from the group consisting of yeast, bacteria, plant, and virus.

9. A virus comprising the isolated polynucleotide of Claim 1.

10. A composition consisting of a polypeptide of at least 750 amino acids that has at least 85% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a polypeptide of SEQ ID NOs:2, 4, 6, 8, 10 and 16.

11. A composition comprising an isolated polynucleotide comprising a nucleotide sequence encoding a first polypeptide of at least 45 amino acids that has at least 70% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:12; or

an isolated polynucleotide comprising the complement of the nucleotide sequence.

12. The composition of Claim 11, wherein the isolated polynucleotide sequence consists of a nucleic acid sequence of SEQ ID NO:11 that codes for the polypeptide of SEQ ID NO:12.

13. The composition of Claim 11 wherein the isolated polynucleotide is DNA.

14. The composition of Claim 11 wherein the isolated polynucleotide is RNA.

15. A chimeric gene comprising the isolated polynucleotide of Claim 11 operably linked to suitable regulatory sequences.

16. An isolated host cell comprising the chimeric gene of Claim 15.

17. An isolated host cell comprising an isolated polynucleotide of Claim 11 or Claim 13.

18. The isolated host cell of Claim 17 wherein the isolated host selected from the group consisting of yeast, bacteria, plant, and virus.

19. A virus comprising the isolated polynucleotide of Claim 11.

20. A composition consisting of a polypeptide of at least 45 amino acids that has at least 85% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO: 12.

21. A composition consisting of an isolated polynucleotide comprising a nucleotide sequence encoding a first polypeptide of at least 750 amino acids that has at least 70% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:14; or

an isolated polynucleotide comprising the complement of the nucleotide sequence.

22. The composition of Claim 21, wherein the isolated nucleotide sequence consists of a nucleic acid sequence of SEQ ID NO: 13 that codes for the polypeptide of SEQ ID NO: 14.

23. The composition of Claim 21 wherein the isolated polynucleotide is DNA.

24. The composition of Claim 21 wherein the isolated polynucleotide is RNA.

25. A chimeric gene comprising the isolated polynucleotide of Claim 21 operably linked to suitable regulatory sequences.

26. An isolated host cell comprising the chimeric gene of Claim 25.

27. An isolated host cell comprising an isolated polynucleotide of Claim 21 or Claim 23.

28. The isolated host cell of Claim 27 wherein the isolated host selected from the group consisting of yeast, bacteria, plant, and virus.

29. A virus comprising the isolated polynucleotide of Claim 21.

30. A composition comprising a polypeptide of at least 750 amino acids that has at least 95% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:14.

31. A method of selecting an isolated polynucleotide that affects the level of expression of a raffinose synthase in a host cell, the method comprising the steps of:

(a) constructing an isolated polynucleotide comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13 and 15 and the complement of such nucleotide sequences;

(b) introducing the isolated polynucleotide into a host cell; and

(c) measuring the level of a raffinose synthase in the host cell containing the polynucleotide.

32. A method of selecting an isolated polynucleotide that affects the level of expression of raffinose synthase in a plant cell, the method comprising the steps of:

(a) constructing an isolated polynucleotide of Claim 1, Claim 11 or Claim 12;

(b) introducing the isolated polynucleotide into a plant cell; and

(c) measuring the level of raffinose synthase in the plant cell containing the polynucleotide.

33. A method of obtaining a nucleic acid fragment encoding a raffinose synthase polypeptide comprising the steps of:

(a) synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15 and the complement of such nucleotide sequences; and

(b) amplifying a nucleic acid sequence using the oligonucleotide primer.

34. A method of obtaining a nucleic acid fragment encoding a raffinose synthase polypeptide comprising the steps of:

(a) probing a cDNA or genomic library with an isolated polynucleotide comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13 and 15 and the complement of such nucleotide sequences;

(b) identifying a DNA clone that hybridizes with the isolated polynucleotide; and

(c) isolating the identified DNA clone.

35. An expression cassette comprising an isolated polynucleotide of Claim 1, Claim 11 or Claim 21 operably linked to a promoter.

36. A method of positive selection of a transformed cell comprising:

(a) transforming a plant cell with the expression cassette of Claim 35; and

(b) growing the transformed plant under conditions allowing expression of the polynucleotide in an amount sufficient to reduce raffinose saccharide content of the plant to provide a positive selection means.

37. The method of Claim 36 wherein the plant cell is a dicot.

38. A method of positive selection of a transformed cell comprising:

transforming a plant cell with the chimeric nucleic acid sequence of Claim 5, Claim 15 or Claim 25; and

growing the transformed plant under conditions allowing expression of the polynucleotide in an amount sufficient to reduce raffinose saccharide content of the plant to provide a positive selection means.

39. The method of Claim 38 wherein the plant cell is a dicot.

5

SEQUENCE LISTING

<110> E. I. du Pont de Nemours and Company

<120> Plant Raffinose Synthase Homologs

<130> BB1251

<140>
<141>

<150> 60/105,451
<151> 1998-10-23

<160> 16

<170> Microsoft Office 97

<210> 1
<211> 1816
<212> DNA
<213> Zea mays

<400> 1
gcacgagccc tgtgggaatc gcttgcctat ctgacaactc agccaacttc gcaaacaggc 60
tgactcacat cagggagaac cacaagtttc agaaaaatgg cagggagggt cacagggaag 120
atgacccagc gaagggccta gcacacgtcg tcaatgagat taaggggaag catcagctca 180
agtatgtgta cgtatggcat gccatcaccg gatactgggg cggagtgagg ccgggtgcag 240
ctggaatgga gcactacgga tcaaagatgc agcggcccgt gccatcgccg ggggttcaga 300
agaacgagcg ctgcgacgcc ctggacagca tgacggccaa cgggctgggc ctcgtgaacc 360
ctgacagggc gttcagtttc tacgacgagc tccactcgta cctcgcgtct gccgggatcg 420
acggggtgaa ggtggacgtg cagaacgtcc tcgagacgct gggcgccggc catggcggga 480
gggtgatgct ggcgaggaag taccagcagg ctctggaagc gtccgtcgcc aggaacttcc 540
ctgacaatgg catcatatcg tgcatgagcc acagcacgga caacttgtac agctcgaaac 600
ggagcgcggt gattagagcc tctgacgatt tctggccgag agaccccgct tcccatacca 660
tacacgtcgc gtccgtcgct tataacaccg tctttcttgg ggagttcatg cagccagact 720
gggacatgtt ccatagtgtt catcccatgg ctgagtacca tgctgcggct cgagcggttg 780
gtggctgtgc catatatgtc agcgacaagc ctgggagcca tgacttcaat ctgctcaaga 840
agctcgtgct tcccgacgga tcgatcctgc gcgcaaagct ccccgggagg ccaaccagag 900
actgcctctt ctctgacccc gcaagggacg gcaaaagcgt tctcaagata tggaacctga 960
acgaacactc cggcgtcgtt ggcgccttca actgccaagg cgccggctgg tgccgggtag 1020
ccaagaagaa cctcatccac gaccagcagc ccggaacggt gagcggcgtc atccgggcgc 1080
aggacgtgga gcaccttgga agggtggctg atcacggctg gaacggcgac gtggtcgtgt 1140
atttgcacgt gggaggggag gtggtgtacc tgccgaagaa cgccttgctg cctgtgacgc 1200
tgagatcgcg cgagtatgag gtgttcaccg tcgtccctct caagcacctg ccaaacggta 1260
cctcctttgc ggcgatcggc cttctcggca tgttcaactc cggtggcgcg gtgagggagc 1320
tgagattcgg tggtgaggat gccgacgtcg agctcagagt gcggggctcg ggcacggtcg 1380
gagcttattc ctcgaccaag ccaacgtgtg tcgccgtcga ttccaaggcg gttggtttct 1440
cctacgatgc cacctgtggc ctcatcagct tcgagctcgg cattcccgac caagaaatgt 1500
acttgtggac ggttacagta gagtattgaa gttacaactg agaattgaga tagattttct 1560
aatgcgtcag agaacgcccg gtggtgtagg tattctgatt tattcagtgt tgtactttct 1620
attgttagtt atatatattc tgtacacatc ccgggatttt atgacaaatc tagataagtg 1680
gcggtcctta gattttagga gcccagaacg aaattttaaa tagaagccct tgttaatatt 1740
gtatattgat tttacaaata aaaatcgaat gtgatcaatt aaaaatcaca accttgataa 1800
aaaaaaaaaa aaaaaa 1816

<210> 2
<211> 508
<212> PRT
<213> Zea mays

<400> 2

```
Thr Ser Pro Val Gly Ile Ala Cys Leu Ser Asp Asn Ser Ala Asn Phe
 1               5                  10                  15

Ala Asn Arg Leu Thr His Ile Arg Glu Asn His Lys Phe Gln Lys Asn
            20                  25                  30

Gly Arg Glu Gly His Arg Glu Asp Asp Pro Ala Lys Gly Leu Ala His
        35                  40                  45

Val Val Asn Glu Ile Lys Gly Lys His Gln Leu Lys Tyr Val Tyr Val
    50                  55                  60

Trp His Ala Ile Thr Gly Tyr Trp Gly Gly Val Arg Pro Gly Ala Ala
65                  70                  75                  80

Gly Met Glu His Tyr Gly Ser Lys Met Gln Arg Pro Val Pro Ser Pro
                85                  90                  95

Gly Val Gln Lys Asn Glu Arg Cys Asp Ala Leu Asp Ser Met Thr Ala
            100                 105                 110

Asn Gly Leu Gly Leu Val Asn Pro Asp Arg Ala Phe Ser Phe Tyr Asp
        115                 120                 125

Glu Leu His Ser Tyr Leu Ala Ser Ala Gly Ile Asp Gly Val Lys Val
    130                 135                 140

Asp Val Gln Asn Val Leu Glu Thr Leu Gly Ala Gly His Gly Gly Arg
145                 150                 155                 160

Val Met Leu Ala Arg Lys Tyr Gln Gln Ala Leu Glu Ala Ser Val Ala
                165                 170                 175

Arg Asn Phe Pro Asp Asn Gly Ile Ile Ser Cys Met Ser His Ser Thr
            180                 185                 190

Asp Asn Leu Tyr Ser Ser Lys Arg Ser Ala Val Ile Arg Ala Ser Asp
        195                 200                 205

Asp Phe Trp Pro Arg Asp Pro Ala Ser His Thr Ile His Val Ala Ser
    210                 215                 220

Val Ala Tyr Asn Thr Val Phe Leu Gly Glu Phe Met Gln Pro Asp Trp
225                 230                 235                 240

Asp Met Phe His Ser Val His Pro Met Ala Glu Tyr His Ala Ala Ala
                245                 250                 255

Arg Ala Val Gly Gly Cys Ala Ile Tyr Val Ser Asp Lys Pro Gly Ser
            260                 265                 270

His Asp Phe Asn Leu Leu Lys Lys Leu Val Leu Pro Asp Gly Ser Ile
        275                 280                 285

Leu Arg Ala Lys Leu Pro Gly Arg Pro Thr Arg Asp Cys Leu Phe Ser
    290                 295                 300
```

```
Asp Pro Ala Arg Asp Gly Lys Ser Val Leu Lys Ile Trp Asn Leu Asn
305                 310                 315                 320

Glu His Ser Gly Val Val Gly Ala Phe Asn Cys Gln Gly Ala Gly Trp
                325                 330                 335

Cys Arg Val Ala Lys Lys Asn Leu Ile His Asp Gln Gln Pro Gly Thr
            340                 345                 350

Val Ser Gly Val Ile Arg Ala Gln Asp Val Glu His Leu Gly Arg Val
        355                 360                 365

Ala Asp His Gly Trp Asn Gly Asp Val Val Val Tyr Leu His Val Gly
    370                 375                 380

Gly Glu Val Val Tyr Leu Pro Lys Asn Ala Leu Leu Pro Val Thr Leu
385                 390                 395                 400

Arg Ser Arg Glu Tyr Glu Val Phe Thr Val Val Pro Leu Lys His Leu
                405                 410                 415

Pro Asn Gly Thr Ser Phe Ala Ala Ile Gly Leu Leu Gly Met Phe Asn
            420                 425                 430

Ser Gly Gly Ala Val Arg Glu Leu Arg Phe Gly Gly Glu Asp Ala Asp
        435                 440                 445

Val Glu Leu Arg Val Arg Gly Ser Gly Thr Val Gly Ala Tyr Ser Ser
    450                 455                 460

Thr Lys Pro Thr Cys Val Ala Val Asp Ser Lys Ala Val Gly Phe Ser
465                 470                 475                 480

Tyr Asp Ala Thr Cys Gly Leu Ile Ser Phe Glu Leu Gly Ile Pro Asp
                485                 490                 495

Gln Glu Met Tyr Leu Trp Thr Val Thr Val Glu Tyr
            500                 505
```

<210> 3
<211> 2868
<212> DNA
<213> Zea mays

<400> 3

```
gcacgagcgc agtccgagtc cagggggcag gggcaccatt accgagtagt ttaattagtg   60
ctagttgtat taccgttgtt agtgacataa ttaccctcac gcgaagaacc cctaccttgg  120
ctagtcctcc tacgtatacg tctcttctct tcgcttaatc ttggtcggtg atttgattct  180
gcacttcttg ggcgcattgc agagggtgct cggtctgcag ctatcttagg ttaaccagtc  240
gtgtgtggtc tatcggtttc gatatagtac tactgctctg ttataatcag cggggaggag  300
gaagatgacg gtgacgccac ggatcacggt gagcgacggg cggctgacgg tgcgcggccg  360
tacggtgctc accggcgtgc cggacaacgt gtcggcggcg cacgcggccg gggcggggct  420
cgtcgacggg gccttcgtcg gcgcccacgc cggcgaggcc aagagccacc acgtcttcac  480
cttcggcacg ctccgggact gccggttcct gtgcctgttc cggttcaagc tgtggtggat  540
gacgcagcgg atgggcgtct ccggccgcga cgtccctctg gagacccagt tcatgctcgt  600
cgaggtccct gccagtgacg gcgacggcga cgacgccccc gcgtacgtgg tgatgcttcc  660
gctgctggag gggcagtttc gggcggcgct gcaggggaac gaccgcgacg agctgcagat  720
ctgcatcgag agcggggaca aggcggtgca gacggaccag gccgcgcaca tggtgtacct  780
ccacgccggc gacaacccct tcgacaccgt caccgccgcc gtcaaggcgg tggagaagca  840
cctgcagacg ttccaccacc gcgacaagaa gaagctgccg tcgttcctcg actggttcgg  900
```

```
ctggtgcacc tgggacgcct tctacaccga cgtcaccgcc gacggcgtca agaacggcct  960
tcagagcctg tccaagggcg gcgcgccgcc gcggttcctc atcatcgacg acggctggca 1020
gcagatcgcc tccgagaaca agcccgaccc caacgtcgcc gtccaggagg gcgcgcaatt 1080
cgccagccgg ctgaccggga tcaaggagaa caccaagttc cagaccaagc ccgacggaga 1140
cggcgacggc gagcaggcgg cgggcgggct gaagcgactg gtggccgaga ccaaggacgc 1200
gcacggcgtg aagcaggtgt acgtgtggca cgccatggcc gggtactggg gcggcgtgac 1260
gccgacggcg gggacggcga tggagcgcta cgagccggcg ctggcgtacc ccgtgcagtc 1320
cccgggcgtg acgggcaacc agccggacat cgtcatggac tcgctgtccg tcctggggct 1380
gggcctggtg cacccgcgcc gggtgcggga cttctacggc gagctccacg cgtacctcgc 1440
ctcctgcggc gtcgacggcg tcaaggtgga cgtgcagaac atcatcgaga cgctgggcgc 1500
cggccacggc ggccgcgtcg ccatcacccg cgcctaccac cgcgcgctcg aggcctccgt 1560
ggcgcgcagc ttcccagaca acggctgcat ctcctgcatg tgccacaact ccgacatgct 1620
gtacagcgcc aggcagaccg ccgtcgtgcg cgcctccgac gacttctacc cgcgcgaccc 1680
ggcatcgcac accgtccacg tcgcctccgt cgcgtacaac accgtcttcc tcggcgagtt 1740
catgcagccc gattgggaca tgttccatag cttgcatccg gcggcggagt accacggcgc 1800
ggcgagggcc atcggtggct gcccgatata cgtcagcgac aagccgggga accacaactt 1860
cgagctgctc aggaagctcg tgctccccga cggctccgtg ctacgcgcgc agcttcccgg 1920
ccggcccaca cgcgactgcc tcttctccga cccggcgcgc gacggcgaga gtttgctcaa 1980
gatctggaac ctgaacaagt gcggtggcgt ggtgggtgtg ttcaactgcc agggagccgg 2040
gtggtgccgc gtgaccaagc ggacgcgcgt gcacgacgcg tcgccgggca cgctgaccgg 2100
caccgtgcgt gccgacgacg tcgacgccat agcgcgcata gctggtgacg gcggcgggtg 2160
ggacggcgag accgtggtgt atgcgcaccg cacgcgggag ctagtgcgac tgccgcgggg 2220
cgtcgcgctg cccgtgacgc taggtccgct ccagtatgag gtgttccatg tgtgcccgct 2280
ccgcgccgtc gtgccggggg tctcgttcgc gcccgtcggg ctgctggaca tgttcaacgc 2340
cgggggcgcc gttgaggagt gcgacgtgat cagcgatgtc ggcggcaagg ccatggctct 2400
cagggttcgc gggtgcggtc ggttcggcgc ttactgctcg cgggagccgg cgaggtgcct 2460
attggactcg gcggaagtgg agttcagcta cgattacgac actggcctcg tgtccgtcga 2520
cttgcgtgtg ccggagcagg agctatacct ttggacgctg gagattatga tctaggtcgt 2580
gatgctgtgc gccgacgcca acgcttgcct tcgcggctga gccattatga gtcggacgac 2640
attgtcattt gctatgtgca cgagagatag atgtattgtg gtggtgttgt ataattggca 2700
tggggagtgt ttttttttgtt atgttcttat ctgtatatca tacgtgctct gaatgttata 2760
taatacttga taataaagga tggtgctcct tttaaaaaaa aaaaaaaaaa aaaaaaaaaa 2820
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa              2868
```

<210> 4
<211> 756
<212> PRT
<213> Zea mays

<400> 4

```
Met Thr Val Thr Pro Arg Ile Thr Val Ser Asp Gly Arg Leu Thr Val
  1               5                  10                  15

Arg Gly Arg Thr Val Leu Thr Gly Val Pro Asp Asn Val Ser Ala Ala
             20                  25                  30

His Ala Ala Gly Ala Gly Leu Val Asp Gly Ala Phe Val Gly Ala His
         35                  40                  45

Ala Gly Glu Ala Lys Ser His His Val Phe Thr Phe Gly Thr Leu Arg
     50                  55                  60

Asp Cys Arg Phe Leu Cys Leu Phe Arg Phe Lys Leu Trp Trp Met Thr
 65                  70                  75                  80

Gln Arg Met Gly Val Ser Gly Arg Asp Val Pro Leu Glu Thr Gln Phe
                 85                  90                  95

Met Leu Val Glu Val Pro Ala Ser Asp Gly Asp Gly Asp Asp Ala Pro
            100                 105                 110
```

```
Ala Tyr Val Val Met Leu Pro Leu Leu Glu Gly Gln Phe Arg Ala Ala
        115                 120                 125

Leu Gln Gly Asn Asp Arg Asp Glu Leu Gln Ile Cys Ile Glu Ser Gly
    130                 135                 140

Asp Lys Ala Val Gln Thr Asp Gln Ala Ala His Met Val Tyr Leu His
145                 150                 155                 160

Ala Gly Asp Asn Pro Phe Asp Thr Val Thr Ala Ala Val Lys Ala Val
                165                 170                 175

Glu Lys His Leu Gln Thr Phe His His Arg Asp Lys Lys Lys Leu Pro
            180                 185                 190

Ser Phe Leu Asp Trp Phe Gly Trp Cys Thr Trp Asp Ala Phe Tyr Thr
        195                 200                 205

Asp Val Thr Ala Asp Gly Val Lys Asn Gly Leu Gln Ser Leu Ser Lys
    210                 215                 220

Gly Gly Ala Pro Pro Arg Phe Leu Ile Ile Asp Asp Gly Trp Gln Gln
225                 230                 235                 240

Ile Ala Ser Glu Asn Lys Pro Asp Pro Asn Val Ala Val Gln Glu Gly
                245                 250                 255

Ala Gln Phe Ala Ser Arg Leu Thr Gly Ile Lys Glu Asn Thr Lys Phe
            260                 265                 270

Gln Thr Lys Pro Asp Gly Asp Gly Asp Gly Glu Gln Ala Ala Gly Gly
        275                 280                 285

Leu Lys Arg Leu Val Ala Glu Thr Lys Asp Ala His Gly Val Lys Gln
    290                 295                 300

Val Tyr Val Trp His Ala Met Ala Gly Tyr Trp Gly Gly Val Thr Pro
305                 310                 315                 320

Thr Ala Gly Thr Ala Met Glu Arg Tyr Glu Pro Ala Leu Ala Tyr Pro
                325                 330                 335

Val Gln Ser Pro Gly Val Thr Gly Asn Gln Pro Asp Ile Val Met Asp
            340                 345                 350

Ser Leu Ser Val Leu Gly Leu Gly Leu Val His Pro Arg Arg Val Arg
        355                 360                 365

Asp Phe Tyr Gly Glu Leu His Ala Tyr Leu Ala Ser Cys Gly Val Asp
    370                 375                 380

Gly Val Lys Val Asp Val Gln Asn Ile Ile Glu Thr Leu Gly Ala Gly
385                 390                 395                 400

His Gly Gly Arg Val Ala Ile Thr Arg Ala Tyr His Arg Ala Leu Glu
                405                 410                 415

Ala Ser Val Ala Arg Ser Phe Pro Asp Asn Gly Cys Ile Ser Cys Met
            420                 425                 430
```

```
Cys His Asn Ser Asp Met Leu Tyr Ser Ala Arg Gln Thr Ala Val Val
        435                 440                 445

Arg Ala Ser Asp Asp Phe Tyr Pro Arg Asp Pro Ala Ser His Thr Val
    450                 455                 460

His Val Ala Ser Val Ala Tyr Asn Thr Val Phe Leu Gly Glu Phe Met
465                 470                 475                 480

Gln Pro Asp Trp Asp Met Phe His Ser Leu His Pro Ala Ala Glu Tyr
                485                 490                 495

His Gly Ala Ala Arg Ala Ile Gly Gly Cys Pro Ile Tyr Val Ser Asp
            500                 505                 510

Lys Pro Gly Asn His Asn Phe Glu Leu Leu Arg Lys Leu Val Leu Pro
        515                 520                 525

Asp Gly Ser Val Leu Arg Ala Gln Leu Pro Gly Arg Pro Thr Arg Asp
    530                 535                 540

Cys Leu Phe Ser Asp Pro Ala Arg Asp Gly Glu Ser Leu Leu Lys Ile
545                 550                 555                 560

Trp Asn Leu Asn Lys Cys Gly Gly Val Val Gly Val Phe Asn Cys Gln
                565                 570                 575

Gly Ala Gly Trp Cys Arg Val Thr Lys Arg Thr Arg Val His Asp Ala
            580                 585                 590

Ser Pro Gly Thr Leu Thr Gly Thr Val Arg Ala Asp Asp Val Asp Ala
        595                 600                 605

Ile Ala Arg Ile Ala Gly Asp Gly Gly Gly Trp Asp Gly Glu Thr Val
    610                 615                 620

Val Tyr Ala His Arg Thr Arg Glu Leu Val Arg Leu Pro Arg Gly Val
625                 630                 635                 640

Ala Leu Pro Val Thr Leu Gly Pro Leu Gln Tyr Glu Val Phe His Val
                645                 650                 655

Cys Pro Leu Arg Ala Val Val Pro Gly Val Ser Phe Ala Pro Val Gly
            660                 665                 670

Leu Leu Asp Met Phe Asn Ala Gly Gly Ala Val Glu Glu Cys Asp Val
        675                 680                 685

Ile Ser Asp Val Gly Gly Lys Ala Met Ala Leu Arg Val Arg Gly Cys
    690                 695                 700

Gly Arg Phe Gly Ala Tyr Cys Ser Arg Glu Pro Ala Arg Cys Leu Leu
705                 710                 715                 720

Asp Ser Ala Glu Val Glu Phe Ser Tyr Asp Tyr Asp Thr Gly Leu Val
                725                 730                 735

Ser Val Asp Leu Arg Val Pro Glu Gln Glu Leu Tyr Leu Trp Thr Leu
            740                 745                 750
```

Glu Ile Met Ile
755

<210> 5
<211> 3060
<212> DNA
<213> Oryza sativa

<220>
<221> unsure
<222> (798) ... (827)

<400> 5
gcacgagggc atacaccacc actcacctca ctcacgtgac gactcgtcca acccaaacag 60
ctcaacccga ccagccaacc caatcctgca cacgcacgcg gcacgcgcgc agtgtcagag 120
tcacaggcca gtcaccaacc aacccgatcc gatccagccg ccgccgccgc cgccaatcac 180
ccccgccaat gtccgtgccg gtggcgcgaa ggctgctcat gatctccacc acccgcgcca 240
cgccgcccgc ctcctcctct ccccttctcc cgcgcccgcg caccatatat accttccctt 300
cccgcgcatt tccccccatc gcatccgccc ctctcccctc ctcttctcgc cggtgccaga 360
ggcggagcag gtgcgccgcg tcgtcgtcgg agatgacggt cacgtcgtcg gtgaaggtcg 420
ctggtgggga gctctcggtc catgggcgga cggtgctgtc cggggtgccg gaggcggtgc 480
gtgcgtcgtc cgccgcggcg gcggggcccg tcgatggggt cttcctcggc ggagacttcg 540
ccgagccggc atcccgccac gtcgtctccc tcggcgcgat gaggggatg cgattcatgg 600
agtgcttccg gttcaagctg tggtggatgg cgcagaggat gggggagaag ggcggcgacg 660
tgccgcacga gacgcagttc ctgctggtgg agtccaagac cggagtcgat ggcgagcagg 720
cggcttgaac agcgtcggta cagcgtgcct ctgttcactt cgtagcgccc gcgctgaagg 780
atgatcctcc gccgggcnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnctt gttcttcctc 840
ccgcttgttc ggagggcgcg ttccggccca gcctccaggg gggcggcgcg ggcggcgacg 900
agctccagct ctgcgtcgag agcggcgacg cgggcacgcg cgccgcttcg ttcgaccgcg 960
cgcttttcgt cggccccgcg gactccgacc ccttcgccgc catcgccggc gccgtcgccg 1020
ccgccaagtc ctgcctcaag acgttccgca tccgcgccga gaagaagctc ccgggcatcg 1080
tcgactactt cgggtggtgc acctgggacg ccttctacca ggacgtcacc caggagggcg 1140
tcgaggccgg cctccgcagc ctcaccgccg gtggcgcgcc gcccaagttc gtcatcatcg 1200
acgacggctg gcagtccgtc ggcaccgacc accagaaccc cgacgacacc ggcgccgacg 1260
ccaaggacaa gcagccgctc ctcgcgcggc tcaccggaat caaggagaac agcaagttcc 1320
aggacggcga cgacccggcc gccggcatca agacggtggt gcgcgccgcg aaggagaagt 1380
acggcctcaa gtacgtctac gtctggcacg ccatcaccgg ctactggggc ggcgtccgcc 1440
ccggcgtcgc ggggatggag gggtaccact ccaacatgca gttccccaac gtctcgccgg 1500
gcgtcgtcga gaacgagccc ggcatgaaga ccgacgtgct caccacccag gggctcgggc 1560
tcgtgcaccc gcgcgccgtg taccgcttct acgacgagct ccacgcctac ctcgccgccg 1620
ccggcgtcga cggcgtcaag gtcgacgtgc agtgcatcct ggagacgctc ggcgccggcc 1680
acggcggccg cgtctcgctc acccgccagt tccaccaggc cctcgacgcc tccattgcca 1740
agaacttccc tgagaacggc atcatcgcct gcatgagcca ccacaccgac gccctctact 1800
gtgctaagca gacggcagtg gtgagagcat cggatgattt ctacccgagg gatccggtgt 1860
ctcacacgat ccacattgcc tcggtggcgt acaacagcgt attcctcggc gagttcatgc 1920
tcccggattg ggacatgttc cactctcttc atcctgccgg cgactaccac ggatcggcga 1980
gggcgatcag cggcggccct gtctatgtca gtgatgcccc tgggaagcac aactttgagc 2040
tgctgaagaa gatggtcttg cccgatggct ccgttcttcg cgcgtggcta cctggccggc 2100
caaccaagga ctgcctcttc accgatccgg cgcgtgacgg cgtcagcttg ctgaagattt 2160
ggaacatgaa caagttcacc ggagtgctcg gagtctacaa ttgccagggc gcggcatgga 2220
gctctgtgga gaagaagaac atcttccaca agactggcgc cgaggccctt tcttgtggcg 2280
tcaagggcag tgacgtccat cttatcgccg acgccgcgac agattccgag tggaacggtg 2340
attgtgcggt gtaccgtcat gccagtgctg atcttgtagt ccttccaaat ggcgcagcac 2400
tgcccatctc cctcaaggtc ttggaacatg acatcctcac cgtgtcccca atcaaggatt 2460
tggcacccgg attcaggttt gcaccaatcg ggctagtcga catgttcaac agcggggcag 2520
cggtcgaagg cctgacctat catcgtctcg acggcgtaaa gtcactcagc aatggctctg 2580
cttccacctt gcctgagctt cagagcttga gctcccaggc catcggattg gtctgcatgg 2640
aagtgagggg atgtggaaag tttggtgcct actcttcagt caggccaaga aagtgcatgc 2700
taggctcagc tcaggtggag ttcacctatg attcctcatc tgggctggtg atccttgacc 2760

tggagaccat gcccaaggaa agggttcaca agattgttgt tgagttgtaa gattcattgt 2820
agtggtatca ctgtatcaca aacacatcag aatgtgagct cttgtgagtc cattatgatg 2880
ttcaggagca gcatatgttg gttgtaagaa tttggagagg tataggattt gatagtgcac 2940
atatctgatt cttggctcaa ttgaaggatg gaatatatgg tgacattttt tctttcagaa 3000
gaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 3060

<210> 6
<211> 770
<212> PRT
<213> Oryza sativa

<220>
<221> UNSURE
<222> (101) ... (110)

<400> 6
Met Thr Val Thr Ser Ser Val Lys Val Ala Gly Gly Glu Leu Ser Val
1 5 10 15

His Gly Arg Thr Val Leu Ser Gly Val Pro Glu Ala Val Arg Ala Ser
20 25 30

Ser Ala Ala Ala Ala Gly Pro Val Asp Gly Val Phe Leu Gly Gly Asp
35 40 45

Phe Ala Glu Pro Ala Ser Arg His Val Val Ser Leu Gly Ala Met Arg
50 55 60

Gly Met Arg Phe Met Glu Cys Phe Arg Phe Lys Leu Trp Trp Met Ala
65 70 75 80

Gln Arg Met Gly Glu Lys Gly Gly Asp Val Pro His Glu Thr Gln Phe
85 90 95

Leu Leu Val Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Leu Val
100 105 110

Leu Pro Pro Ala Cys Ser Glu Gly Ala Phe Arg Pro Ser Leu Gln Gly
115 120 125

Gly Gly Ala Gly Gly Asp Glu Leu Gln Leu Cys Val Glu Ser Gly Asp
130 135 140

Ala Gly Thr Arg Ala Ala Ser Phe Asp Arg Ala Leu Phe Val Gly Pro
145 150 155 160

Ala Asp Ser Asp Pro Phe Ala Ala Ile Ala Gly Ala Val Ala Ala Ala
165 170 175

Lys Ser Cys Leu Lys Thr Phe Arg Ile Arg Ala Glu Lys Lys Leu Pro
180 185 190

Gly Ile Val Asp Tyr Phe Gly Trp Cys Thr Trp Asp Ala Phe Tyr Gln
195 200 205

Asp Val Thr Gln Glu Gly Val Glu Ala Gly Leu Arg Ser Leu Thr Ala
210 215 220

Gly Gly Ala Pro Pro Lys Phe Val Ile Ile Asp Asp Gly Trp Gln Ser
225 230 235 240

Val Gly Thr Asp His Gln Asn Pro Asp Asp Thr Gly Ala Asp Ala Lys
245 250 255

Asp Lys Gln Pro Leu Leu Ala Arg Leu Thr Gly Ile Lys Glu Asn Ser
260 265 270

Lys Phe Gln Asp Gly Asp Asp Pro Ala Ala Gly Ile Lys Thr Val Val
275 280 285

Arg Ala Ala Lys Glu Lys Tyr Gly Leu Lys Tyr Val Tyr Val Trp His
290 295 300

Ala Ile Thr Gly Tyr Trp Gly Gly Val Arg Pro Gly Val Ala Gly Met
305 310 315 320

Glu Gly Tyr His Ser Asn Met Gln Phe Pro Asn Val Ser Pro Gly Val
325 330 335

Val Glu Asn Glu Pro Gly Met Lys Thr Asp Val Leu Thr Thr Gln Gly
340 345 350

Leu Gly Leu Val His Pro Arg Ala Val Tyr Arg Phe Tyr Asp Glu Leu
355 360 365

His Ala Tyr Leu Ala Ala Ala Gly Val Asp Gly Val Lys Val Asp Val
370 375 380

Gln Cys Ile Leu Glu Thr Leu Gly Ala Gly His Gly Gly Arg Val Ser
385 390 395 400

Leu Thr Arg Gln Phe His Gln Ala Leu Asp Ala Ser Ile Ala Lys Asn
405 410 415

Phe Pro Glu Asn Gly Ile Ile Ala Cys Met Ser His His Thr Asp Ala
420 425 430

Leu Tyr Cys Ala Lys Gln Thr Ala Val Val Arg Ala Ser Asp Asp Phe
435 440 445

Tyr Pro Arg Asp Pro Val Ser His Thr Ile His Ile Ala Ser Val Ala
450 455 460

Tyr Asn Ser Val Phe Leu Gly Glu Phe Met Leu Pro Asp Trp Asp Met
465 470 475 480

Phe His Ser Leu His Pro Ala Gly Asp Tyr His Gly Ser Ala Arg Ala
485 490 495

Ile Ser Gly Gly Pro Val Tyr Val Ser Asp Ala Pro Gly Lys His Asn
500 505 510

Phe Glu Leu Leu Lys Lys Met Val Leu Pro Asp Gly Ser Val Leu Arg
515 520 525

Ala Trp Leu Pro Gly Arg Pro Thr Lys Asp Cys Leu Phe Thr Asp Pro
530 535 540

Ala Arg Asp Gly Val Ser Leu Leu Lys Ile Trp Asn Met Asn Lys Phe
545 550 555 560

```
Thr Gly Val Leu Gly Val Tyr Asn Cys Gln Gly Ala Ala Trp Ser Ser
                565                 570                 575

Val Glu Lys Lys Asn Ile Phe His Lys Thr Gly Ala Glu Ala Leu Ser
            580                 585                 590

Cys Gly Val Lys Gly Ser Asp Val His Leu Ile Ala Asp Ala Ala Thr
        595                 600                 605

Asp Ser Glu Trp Asn Gly Asp Cys Ala Val Tyr Arg His Ala Ser Ala
    610                 615                 620

Asp Leu Val Val Leu Pro Asn Gly Ala Ala Leu Pro Ile Ser Leu Lys
625                 630                 635                 640

Val Leu Glu His Asp Ile Leu Thr Val Ser Pro Ile Lys Asp Leu Ala
                645                 650                 655

Pro Gly Phe Arg Phe Ala Pro Ile Gly Leu Val Asp Met Phe Asn Ser
            660                 665                 670

Gly Ala Ala Val Glu Gly Leu Thr Tyr His Arg Leu Asp Gly Val Lys
        675                 680                 685

Ser Leu Ser Asn Gly Ser Ala Ser Thr Leu Pro Glu Leu Gln Ser Leu
    690                 695                 700

Ser Ser Gln Ala Ile Gly Leu Val Cys Met Glu Val Arg Gly Cys Gly
705                 710                 715                 720

Lys Phe Gly Ala Tyr Ser Ser Val Arg Pro Arg Lys Cys Met Leu Gly
                725                 730                 735

Ser Ala Gln Val Glu Phe Thr Tyr Asp Ser Ser Ser Gly Leu Val Ile
            740                 745                 750

Leu Asp Leu Glu Thr Met Pro Lys Glu Arg Val His Lys Ile Val Val
        755                 760                 765

Glu Leu
    770
```

<210> 7
<211> 2842
<212> DNA
<213> Oryza sativa

<400> 7

```
gcacgagtac agcccgccat ttagttaatt agtagttgtt aattatcttc actaacctta   60
attaccctca cgcaaagcac tctctaatcc tcctcggcgc tcgcataatc ttgtatccct  120
catctcagat tatcatcacc tcgccgtgtt agaaaaaagg gagtaagttc gttttatcgt  180
cggatttggt gagagttcgc cggagtagaa gaagatgacg gtgacgccgc agatcacggt  240
gagcgacggg aggctggcgg tgcgggggcg gacggtgctg accggcgtgc cggagaacgt  300
gacggcggcg cacgcgtccg gcgccgggct cgtcgacggc gccttcgtcg gcgccgacgc  360
cggcgaggcc aagagccacc acgtcttcac cttcggcacg ctccgggaat gccggttcat  420
gtgcctgttc aggttcaagc tgtggtggat gacgcagagg atgggctcct ccggccgcga  480
cgtcccgctc gagacccagt tcatgctcat cgaggtcccc gccaccgccg ccggcgacgg  540
ccacgacggc ggcggcgacg gcgagccggt gttcgtggtg atgctgccgc tgctggaggg  600
gaagttccgc gcggcgctgc agggaaacga cgacgacgag ctccagatat gcatcgagag  660
```

```
cggggataag gcggtgcaga cagagcaggg cgtgaacatg gtgtacatcc acgctgggac  720
caacccttc gacacgatca cccaagccat caaggcagtg gagaagcgaa tgcagacgtt  780
ccaccacagg gacaagaaga agatgccgtc cttcttggat tggttcgggt ggtgcacatg  840
ggacgcgttc tacaccgacg tcaccgccga cggcgtcaag cagggcctcc gcagcctcgc  900
caatggcggc gcgccgccgc gcttcctcat catcgacgac ggctggcagc agatcggcac  960
cgaggacgac gacacggacg aacaccccgc cgtcgccgtc caggagggcg cccagttcgc 1020
cagccggctc accggcatca aggagaacgt caagttccag agcaagaacg gcggcgccgg 1080
cgaggacacg ccggggctga ggatgctggt ggaggaggtg aaggggagc acggcgtccg 1140
gcaggtgtac gtgtggcacg ccatggccgg gtactggggc ggcgtggcgc cggcgccggc 1200
gatggagcgg tacgaggcgg cgctggcgta cccggtgcag tcgccggggg tgacggcgaa 1260
ccagccggac atcgtcatgg actccctctc cgtgctcggc ctgggcctcg tccacccgcg 1320
caaggtgctc gacttctacg acgagctcca cgcctacctg gcgtcgtgcg gcgtcgacgg 1380
cgtgaaggtg gacgtgcaga acatcatcga gacgctcggc gccggccacg gcggccgcgt 1440
cgcgctcacg cgcgcctaca accgcgcgct cgaggcgtcc gtggcgcgga gcttcccgga 1500
caacggctgc atctcctgca tgtgccacaa caccgacatg ctatacagcg cgcgccagac 1560
cgccgtcgtc cgcgcctccg acgacttcta ccctcgcgac ccggcgtccc acaccatcca 1620
cgtcgcctcc gtcgcctaca acaccgtctt cctcggcgag ttcatgcagc ccgactggga 1680
catgttccat agcttgcacc cggcggcgga gtaccacggc gcggcgaggg cgatcggcgg 1740
gtgtccgatc tacgtcagcg acaagccggg gaaccacaac ttcgatctgc tcaggaagct 1800
cgtcctcccc gacggctccg tgctccgggc gcgcctcccc ggccgcccca cccgcgactg 1860
cctcttctcc gacccggcgc gcgacggcga gagtttgctc aagatatgga acctgaacaa 1920
ttgcggcggc gtggtcgggg tgttcaactg ccagggcgcc gggtggtgca gggtggccaa 1980
gaagacgcgc gtgcacgacg cggcgcccgg gacgctcacc ggcgccgtgc gcgccgacga 2040
cgtcgacgcc atcgcgcagg tcgccggcgg cgacggcgga ggctgggacg gggaggccgt 2100
ggtgtacgcg caccgggcgc gtgagctggt ccggctgccg cgcggcgccg cgctgccggt 2160
gacgctcggg gcgctcgagt acgaggtgtt ccacgtctgc cccgtccgcg ccatcgccgc 2220
ggcgccgggc ggcgccgccg tcgcgttcgc gcccgtcggg ctcctcgaca tgttcaacgc 2280
cggcggcgcc gtcgaggagt gcgccgtcga tgccgcagcc gccgtggcgc tcagggtgcg 2340
cggttgcggc cggttcggcg cctacttctc gcggaggccg gcgaggtgcg cgctcgacgg 2400
cgccgacgtg gggttcacct acgacggcga cacggggctc gtcgccgtcg acctgccggt 2460
gccggagcag gagatgtaca ggtggaacct ggagattcac gtctaggatg ctgctgctgc 2520
cacgtggcgg tgcgatcttt gccctctggt gttcgtggct gagacattcc cagccgttgg 2580
atcattgcca ttgatggttt gcagcagaga tggacggacc gatggagaga ttgtgatgtg 2640
ccctcttacg acttgcaacg cggtagaggg agtatgcttt gctacatcta aaatgaaaga 2700
agaaaattat agtatgcttt tgtcatgtta tggcttgtaa aatgcaataa tatgatgtaa 2760
tgctatggtt tatgtattat tattagctga taaataaagg atggtgctcc ttaaaaaaaa 2820
aaaaaaaaaa aaaaaaaaaa aa                                          2842
```

<210> 8
<211> 763
<212> PRT
<213> Oryza sativa

<400> 8

```
Met Thr Val Thr Pro Gln Ile Thr Val Ser Asp Gly Arg Leu Ala Val
  1               5                  10                  15

Arg Gly Arg Thr Val Leu Thr Gly Val Pro Glu Asn Val Thr Ala Ala
             20                  25                  30

His Ala Ser Gly Ala Gly Leu Val Asp Gly Ala Phe Val Gly Ala Asp
         35                  40                  45

Ala Gly Glu Ala Lys Ser His His Val Phe Thr Phe Gly Thr Leu Arg
     50                  55                  60

Glu Cys Arg Phe Met Cys Leu Phe Arg Phe Lys Leu Trp Trp Met Thr
 65                  70                  75                  80
```

```
Gln Arg Met Gly Ser Ser Gly Arg Asp Val Pro Leu Glu Thr Gln Phe
                85                  90                  95

Met Leu Ile Glu Val Pro Ala Thr Ala Ala Gly Asp Gly His Asp Gly
            100                 105                 110

Gly Gly Asp Gly Glu Pro Val Phe Val Val Met Leu Pro Leu Leu Glu
        115                 120                 125

Gly Lys Phe Arg Ala Ala Leu Gln Gly Asn Asp Asp Asp Glu Leu Gln
    130                 135                 140

Ile Cys Ile Glu Ser Gly Asp Lys Ala Val Gln Thr Glu Gln Gly Val
145                 150                 155                 160

Asn Met Val Tyr Ile His Ala Gly Thr Asn Pro Phe Asp Thr Ile Thr
                165                 170                 175

Gln Ala Ile Lys Ala Val Glu Lys Arg Met Gln Thr Phe His His Arg
            180                 185                 190

Asp Lys Lys Lys Met Pro Ser Phe Leu Asp Trp Phe Gly Trp Cys Thr
        195                 200                 205

Trp Asp Ala Phe Tyr Thr Asp Val Thr Ala Asp Gly Val Lys Gln Gly
    210                 215                 220

Leu Arg Ser Leu Ala Asn Gly Gly Ala Pro Pro Arg Phe Leu Ile Ile
225                 230                 235                 240

Asp Asp Gly Trp Gln Gln Ile Gly Thr Glu Asp Asp Asp Thr Asp Glu
                245                 250                 255

His Pro Ala Val Ala Val Gln Glu Gly Ala Gln Phe Ala Ser Arg Leu
            260                 265                 270

Thr Gly Ile Lys Glu Asn Val Lys Phe Gln Ser Lys Asn Gly Gly Ala
        275                 280                 285

Gly Glu Asp Thr Pro Gly Leu Arg Met Leu Val Glu Glu Val Lys Gly
    290                 295                 300

Glu His Gly Val Arg Gln Val Tyr Val Trp His Ala Met Ala Gly Tyr
305                 310                 315                 320

Trp Gly Gly Val Ala Pro Ala Pro Ala Met Glu Arg Tyr Glu Ala Ala
                325                 330                 335

Leu Ala Tyr Pro Val Gln Ser Pro Gly Val Thr Ala Asn Gln Pro Asp
            340                 345                 350

Ile Val Met Asp Ser Leu Ser Val Leu Gly Leu Gly Leu Val His Pro
        355                 360                 365

Arg Lys Val Leu Asp Phe Tyr Asp Glu Leu His Ala Tyr Leu Ala Ser
    370                 375                 380

Cys Gly Val Asp Gly Val Lys Val Asp Val Gln Asn Ile Ile Glu Thr
385                 390                 395                 400
```

Leu Gly Ala Gly His Gly Gly Arg Val Ala Leu Thr Arg Ala Tyr Asn
405 410 415

Arg Ala Leu Glu Ala Ser Val Ala Arg Ser Phe Pro Asp Asn Gly Cys
420 425 430

Ile Ser Cys Met Cys His Asn Thr Asp Met Leu Tyr Ser Ala Arg Gln
435 440 445

Thr Ala Val Val Arg Ala Ser Asp Asp Phe Tyr Pro Arg Asp Pro Ala
450 455 460

Ser His Thr Ile His Val Ala Ser Val Ala Tyr Asn Thr Val Phe Leu
465 470 475 480

Gly Glu Phe Met Gln Pro Asp Trp Asp Met Phe His Ser Leu His Pro
485 490 495

Ala Ala Glu Tyr His Gly Ala Ala Arg Ala Ile Gly Gly Cys Pro Ile
500 505 510

Tyr Val Ser Asp Lys Pro Gly Asn His Asn Phe Asp Leu Leu Arg Lys
515 520 525

Leu Val Leu Pro Asp Gly Ser Val Leu Arg Ala Arg Leu Pro Gly Arg
530 535 540

Pro Thr Arg Asp Cys Leu Phe Ser Asp Pro Ala Arg Asp Gly Glu Ser
545 550 555 560

Leu Leu Lys Ile Trp Asn Leu Asn Asn Cys Gly Gly Val Val Gly Val
565 570 575

Phe Asn Cys Gln Gly Ala Gly Trp Cys Arg Val Ala Lys Lys Thr Arg
580 585 590

Val His Asp Ala Ala Pro Gly Thr Leu Thr Gly Ala Val Arg Ala Asp
595 600 605

Asp Val Asp Ala Ile Ala Gln Val Ala Gly Gly Asp Gly Gly Gly Trp
610 615 620

Asp Gly Glu Ala Val Val Tyr Ala His Arg Ala Arg Glu Leu Val Arg
625 630 635 640

Leu Pro Arg Gly Ala Ala Leu Pro Val Thr Leu Gly Ala Leu Glu Tyr
645 650 655

Glu Val Phe His Val Cys Pro Val Arg Ala Ile Ala Ala Ala Pro Gly
660 665 670

Gly Ala Ala Val Ala Phe Ala Pro Val Gly Leu Leu Asp Met Phe Asn
675 680 685

Ala Gly Gly Ala Val Glu Glu Cys Ala Val Asp Ala Ala Ala Ala Val
690 695 700

Ala Leu Arg Val Arg Gly Cys Gly Arg Phe Gly Ala Tyr Phe Ser Arg
705 710 715 720

Arg Pro Ala Arg Cys Ala Leu Asp Gly Ala Asp Val Gly Phe Thr Tyr
725 730 735

Asp Gly Asp Thr Gly Leu Val Ala Val Asp Leu Pro Val Pro Glu Gln
740 745 750

Glu Met Tyr Arg Trp Asn Leu Glu Ile His Val
755 760

<210> 9
<211> 2524
<212> DNA
<213> Glycine max

<400> 9
gcacgagatt gaatctcaca aaaaatgggt ccaagctcga agaaagcttc agctaaatca 60
ggtgtgacaa agcacatgaa gggcttcagc ctctgcaact caaccctaaa agtaaatggg 120
caagtcatcc tctcccaagt ccccaagaac gtaaccctca ccccatgcac ctacgacact 180
cacaccaccg gatgcttcct cggtttccac gccacctccc caaaatcccg ccacgtggca 240
cccttaggac agcttaaaaa cataagcttc acttccatct tccggttcaa ggtttggtgg 300
accactctct ggaccggctc caacggccgc gacctggaaa ccgaaaccca attcctcatg 360
ctccaatccc acccttatgt tctcttccta cccatcctcc aacccccatt tcgcgcctcg 420
ctgcagcctc actcagacga caacgttgcg gtgtgtgtgg agagcggctc cagccacgta 480
acagcctcat cattcgacac tgtcgtctac ttgcacgcag gggacaaccc cttcacgctg 540
gtcaaggaag ccatgcgcgt cgtccgggcc cacttgggga gcttcaagct tctggaagag 600
aaaacagttc cgggaatggt ggacaagttc ggttggtgca cgtgggacgc cttttacctg 660
acggtgcacc ctgagggcgt cagagagggc gtgaagggcc tggttgacgg cggttgtcct 720
ccgggattcg tcctgatcga cgacggctgg cagtgcatca gccacgattc cgatccggag 780
aaggagggga tgaatcagac ggtggccggg gagcaaatgc cctgcaggtt gattagttac 840
gaggagaatt acaagtttag gagctataag gaagggaagg ggttgaaggg gtttgtgaga 900
gaattgaagg aggagtttgg gtcggtggag tacgtgtacg tgtggcacgc gctgtgcggg 960
tattggggag gggtgaggcc gggggtggcg gggatggcgg aggcggcggt ggagaagcca 1020
aagctgacgg aggggttgaa gggaacgatg gaggatctgg cggtggacaa gattgtgaat 1080
aatggggtcg gggtggtgcc gccggagctg gtgggggaaa tgtatgaggg ccttcacgcg 1140
cacttggaga gtgcgggtat tgatggggtc aaagttgatg tcatccactt gctagaaatg 1200
gtgtgtgaga aatatggagg gcgagtggat atggcgaaag catattacaa agctctcact 1260
gcttccgtga gaaaacattt taagggcaac ggcgtcattg ccagcatgga gcattgcaac 1320
gatttcatgt tgctgggaac tgaagcaata tcccttggtc gtgttgggga tgatttctgg 1380
tgcactgacc cttatggtga tccaaatggt acattttggc tacaagggtg tcacatggtg 1440
cattgtgcat acaacagctt gtggatgggc aatttcatcc acccagattg ggacatgttc 1500
caatctactc atccttgtgc tgccttccat gctgcctcaa gagccatatc tggtggcccc 1560
atttacatca gtgacacagt tgggaaccac aactttgagc tgcttaagac cttggccttg 1620
ccagatgggt ccatcctcag atgtgagcac tatgcactcc caaccaggga ctgtctcttt 1680
gctgaccctc tccatgatgg caaaacaatg ctcaagatat ggaacatcaa caagtacact 1740
ggagttcttg gggtgtttaa ctgccaggga ggaggttggt tccgtgagat taggtccaac 1800
aaatgtgctg ctgagttttc tcatagggta tcaaccaaga ccaatatcaa agacattgaa 1860
tgggatagtg gaaagaatcc aatttccatt gaaggggtgc aacttttcgc ttcgtatttc 1920
agccaagcca agaaactcat cctctcagca ccatctgatg acagtgaaga gatttccttg 1980
gagccattca atttcgagct tataacagtt tcccctgtga ctgtcttgcc tggcaagtca 2040
gtgaagtttg ctcctattgg tttggtgaat atgctaaaca ctggtggagc agtccagtct 2100
ttagcttttg atgagggtca gaatttggtt gaagttggtt taagaggcac tggggagatg 2160
agagtctatg cctcagagaa accaagaacg tgtagaattg atggcaaaga agttgatttt 2220
gaatatgaag ggtctatggt caacattcaa gtaccatggc ctggttcttc aaaattgtcc 2280
actgttcagt atgtatttta agcctagaag tgatttcttt ttaacttttg aatggttgtc 2340
cacagtgaat gatgtgaaag ggctttttct cccttactac cgagtgaaat aatagagtga 2400
tctaataatt gtatcaagag ctagattcct ttttattcaa tgaaagccag tttttttga 2460
gtgttaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2520
aaaa 2524

<210> 10
<211> 758
<212> PRT
<213> Glycine max

<400> 10

```
Met Gly Pro Ser Ser Lys Lys Ala Ser Ala Lys Ser Gly Val Thr Lys
 1               5                  10                  15

His Met Lys Gly Phe Ser Leu Cys Asn Ser Thr Leu Lys Val Asn Gly
            20                  25                  30

Gln Val Ile Leu Ser Gln Val Pro Lys Asn Val Thr Leu Thr Pro Cys
        35                  40                  45

Thr Tyr Asp Thr His Thr Thr Gly Cys Phe Leu Gly Phe His Ala Thr
    50                  55                  60

Ser Pro Lys Ser Arg His Val Ala Pro Leu Gly Gln Leu Lys Asn Ile
65                  70                  75                  80

Ser Phe Thr Ser Ile Phe Arg Phe Lys Val Trp Trp Thr Thr Leu Trp
                85                  90                  95

Thr Gly Ser Asn Gly Arg Asp Leu Glu Thr Glu Thr Gln Phe Leu Met
            100                 105                 110

Leu Gln Ser His Pro Tyr Val Leu Phe Leu Pro Ile Leu Gln Pro Pro
        115                 120                 125

Phe Arg Ala Ser Leu Gln Pro His Ser Asp Asp Asn Val Ala Val Cys
    130                 135                 140

Val Glu Ser Gly Ser Ser His Val Thr Ala Ser Ser Phe Asp Thr Val
145                 150                 155                 160

Val Tyr Leu His Ala Gly Asp Asn Pro Phe Thr Leu Val Lys Glu Ala
                165                 170                 175

Met Arg Val Val Arg Ala His Leu Gly Ser Phe Lys Leu Leu Glu Glu
            180                 185                 190

Lys Thr Val Pro Gly Met Val Asp Lys Phe Gly Trp Cys Thr Trp Asp
        195                 200                 205

Ala Phe Tyr Leu Thr Val His Pro Glu Gly Val Arg Glu Gly Val Lys
    210                 215                 220

Gly Leu Val Asp Gly Gly Cys Pro Pro Gly Phe Val Leu Ile Asp Asp
225                 230                 235                 240

Gly Trp Gln Cys Ile Ser His Asp Ser Asp Pro Glu Lys Glu Gly Met
                245                 250                 255

Asn Gln Thr Val Ala Gly Glu Gln Met Pro Cys Arg Leu Ile Ser Tyr
            260                 265                 270

Glu Glu Asn Tyr Lys Phe Arg Ser Tyr Lys Glu Gly Lys Gly Leu Lys
        275                 280                 285
```

```
Gly Phe Val Arg Glu Leu Lys Glu Glu Phe Gly Ser Val Glu Tyr Val
    290                 295                 300

Tyr Val Trp His Ala Leu Cys Gly Tyr Trp Gly Gly Val Arg Pro Gly
305                 310                 315                 320

Val Ala Gly Met Ala Glu Ala Ala Val Glu Lys Pro Lys Leu Thr Glu
                325                 330                 335

Gly Leu Lys Gly Thr Met Glu Asp Leu Ala Val Asp Lys Ile Val Asn
            340                 345                 350

Asn Gly Val Gly Val Val Pro Pro Glu Leu Val Gly Glu Met Tyr Glu
        355                 360                 365

Gly Leu His Ala His Leu Glu Ser Ala Gly Ile Asp Gly Val Lys Val
    370                 375                 380

Asp Val Ile His Leu Leu Glu Met Val Cys Glu Lys Tyr Gly Gly Arg
385                 390                 395                 400

Val Asp Met Ala Lys Ala Tyr Tyr Lys Ala Leu Thr Ala Ser Val Arg
                405                 410                 415

Lys His Phe Lys Gly Asn Gly Val Ile Ala Ser Met Glu His Cys Asn
            420                 425                 430

Asp Phe Met Leu Leu Gly Thr Glu Ala Ile Ser Leu Gly Arg Val Gly
        435                 440                 445

Asp Asp Phe Trp Cys Thr Asp Pro Tyr Gly Asp Pro Asn Gly Thr Phe
    450                 455                 460

Trp Leu Gln Gly Cys His Met Val His Cys Ala Tyr Asn Ser Leu Trp
465                 470                 475                 480

Met Gly Asn Phe Ile His Pro Asp Trp Asp Met Phe Gln Ser Thr His
                485                 490                 495

Pro Cys Ala Ala Phe His Ala Ala Ser Arg Ala Ile Ser Gly Gly Pro
            500                 505                 510

Ile Tyr Ile Ser Asp Thr Val Gly Asn His Asn Phe Glu Leu Leu Lys
        515                 520                 525

Thr Leu Ala Leu Pro Asp Gly Ser Ile Leu Arg Cys Glu His Tyr Ala
    530                 535                 540

Leu Pro Thr Arg Asp Cys Leu Phe Ala Asp Pro Leu His Asp Gly Lys
545                 550                 555                 560

Thr Met Leu Lys Ile Trp Asn Ile Asn Lys Tyr Thr Gly Val Leu Gly
                565                 570                 575

Val Phe Asn Cys Gln Gly Gly Gly Trp Phe Arg Glu Ile Arg Ser Asn
            580                 585                 590

Lys Cys Ala Ala Glu Phe Ser His Arg Val Ser Thr Lys Thr Asn Ile
        595                 600                 605
```

Lys Asp Ile Glu Trp Asp Ser Gly Lys Asn Pro Ile Ser Ile Glu Gly
610 615 620

Val Gln Leu Phe Ala Ser Tyr Phe Ser Gln Ala Lys Lys Leu Ile Leu
625 630 635 640

Ser Ala Pro Ser Asp Asp Ser Glu Glu Ile Ser Leu Glu Pro Phe Asn
645 650 655

Phe Glu Leu Ile Thr Val Ser Pro Val Thr Val Leu Pro Gly Lys Ser
660 665 670

Val Lys Phe Ala Pro Ile Gly Leu Val Asn Met Leu Asn Thr Gly Gly
675 680 685

Ala Val Gln Ser Leu Ala Phe Asp Glu Gly Gln Asn Leu Val Glu Val
690 695 700

Gly Leu Arg Gly Thr Gly Glu Met Arg Val Tyr Ala Ser Glu Lys Pro
705 710 715 720

Arg Thr Cys Arg Ile Asp Gly Lys Glu Val Asp Phe Glu Tyr Glu Gly
725 730 735

Ser Met Val Asn Ile Gln Val Pro Trp Pro Gly Ser Ser Lys Leu Ser
740 745 750

Thr Val Gln Tyr Val Phe
755

<210> 11
<211> 559
<212> DNA
<213> Glycine max

<220>
<221> unsure
<222> (405)

<220>
<221> unsure
<222> (419)

<220>
<221> unsure
<222> (425)

<220>
<221> unsure
<222> (457)

<220>
<221> unsure
<222> (460)

<220>
<221> unsure
<222> (462)

<220>
<221> unsure
<222> (477)

<220>
<221> unsure
<222> (478)

<220>
<221> unsure
<222> (481)

<220>
<221> unsure
<222> (489)

<220>
<221> unsure
<222> (490)

<220>
<221> unsure
<222> (491)

<220>
<221> unsure
<222> (499)

<220>
<221> unsure
<222> (508)

<220>
<221> unsure
<222> (514)

<220>
<221> unsure
<222> (519)

<220>
<221> unsure
<222> (529)

<220>
<221> unsure
<222> (534)

<220>
<221> unsure
<222> (540)

<220>
<221> unsure
<222> (557)

<400> 11
atttcagaag ctagtgatca atcatcaaca cagagctgag tgtgttgctc tgtgaatcca 60
ccttcttctt ccattggagg accatttcct cctggaatag aaatactacc acacttttct 120
tttttcactt ctctaagttg ctaagttaat tgctccttca ttttttcact cttcgttctc 180

gcgtacccgt gtcgcggtaa ctcgtggtga agtgttcgaa aatgactgtc acacctaaga 240
tctcaagtta acgatgggaa acttgttgtc catggtaaga ccattctgac tggagtgcca 300
agacaacgtt gtgctgactc aagttctgga aggggtcttg tgactggtgc ttttgttggt 360
gccacagctt cacacagcaa aaagtcctca aggtggttca aaatngggtg gttttaaang 420
gggcnccggg ttcaaggtgg ttggtttccg ggttaangtn angggggtgg atccaannat 480
ngggaactnn nggaagggna tttccccngg gaancaatna aggcttatng gaancaaagn 540
aatggaaccg gtggggnga 559

<210> 12
<211> 45
<212> PRT
<213> Glycine max

<400> 12
Val Asn Asp Gly Lys Leu Val Val His Gly Lys Thr Ile Leu Thr Gly
1 5 10 15

Val Pro Arg Gln Arg Cys Ala Asp Ser Ser Ser Gly Arg Gly Leu Val
20 25 30

Thr Gly Ala Phe Val Gly Ala Thr Ala Ser His Ser Lys
35 40 45

<210> 13
<211> 2668
<212> DNA
<213> Triticum aestivum

<400> 13
gcacgaggcg agggcagcgc accatcttat catatcatta gattagtgct actagttccg 60
attacccgca cacaaagcac tctctatcac tcttatcata taccatcgtc cggcatctat 120
caagctccgt atcgaagcaa gaaggggtgg tcctttgaat ctggtcgtcg gagatgacgg 180
tgacaccgca gatcacggtg agcgacggga ggctggcggt gcgcggccgg acggtgctct 240
ccggcgtgcc ggacaacgtt accgcggcgc acgcttccgg ggccgggctc gtcgacgggg 300
ccttcgtcgg cgccacggcc ggcgaggcca agagccacca cgtcttcacc ttcgggactc 360
tccgcgactg ccgattcatg tgcctgttcc ggttcaagct gtggtggatg acgcagcgga 420
tgggcacctc cggccgcgac gtcccgctgg agacccaatt catcctcatc gaggtccctg 480
ccgccgccgg caacgacgac ggcgacagcg agccggtgta cctggtgatg ctgccgctgc 540
tggaggggca gttccgaacg gtgctccagg gcaacgacca agaccagctc cagatctgca 600
tcgagagcgg ggacaaagcg gtgcagacgg agcagggcat gaacagcgtg tacatccacg 660
ccggcaccaa ccccttcgac accatcaccc aggccgtcaa ggccgttgag aagcacatgc 720
agacgttcca ccacagggag aagaaaaagg tgccgtcgtt tgtggactgg ttcgggtggt 780
gcacgtggga cgccttctac acggacgtga cggccgacgg cgtcaagcag gggctccgca 840
gcctggcgga gggtggcgcg ccgccgcggt tcctcatcat tgacgacggc tggcagcaga 900
tcggcagcga gaacaaggaa gacccgagcg tcgccgtcca ggaaggggcg cagttcgcga 960
gcaggctcac cggcatcaag gagaacacca agttccagag cgagcagcag gaggagaccc 1020
cggggctgaa gcggctggtg gaggagacca agaaggagca cggcgtcaag agcgtctacg 1080
tctggcacgc catggccggc tactggggcg gcgtcaagcc gtcggcggcc gggatggagc 1140
actacgagtc cgcgctggcc tacccggtgc agtcgccggg cgtcaccggc aaccagcccg 1200
acatcgtcat ggactcgctc tccgtgctcg gcctcggcct cgtgcacccg cgcaaggtct 1260
acagcttcta cgacgagctc cacgcctacc tggccgcctg cggcgtcgac ggcgtcaagg 1320
tggacgtgca gaacatcgtg gagaccctcg gcgccggcca cggcggccgc gtcgcgctca 1380
cacgcgccta ccaccgcgcg ctcgaggcct ccgtcgcccg caacttcccc gacaacggat 1440
gcatctcctg catgtgccac aacaccgaca tgctctacag cgccaagcag accgccgtcg 1500
tgcgcgcctc cgacgacttc tacccgcgcg acccggcgtc gcacaccgtc cacatctcct 1560
ccgtcgctta caacacgctc ttcctcggcg agttcatgca gcccgactgg gacatgttcc 1620
atagcctgca cccggcggcg gagtaccacg gcgcggcgag ggccatcggc ggctgcccca 1680
tttatgtcag cgacaagcca gggaaccaca acttcgacct tctcaagaag ctggtgctcc 1740
ccgacggctc cgtgctccgc gcacagctcc ccggcaggcc cacgcgcgac tgcctcttct 1800

```
ccgacccggc gcgcgacggt gccagcttgc tcaagatatg gaacatgaac aagtgcgccg 1860
gcgtggtggg ggtgttcaac tgccagggcg cggggtggtg tcgcgtcgtc aaaaagacaa 1920
ggatccacga cgaggcgccc gggacgctca ccggctcggt gcgcgccgag gacgtggagg 1980
gcatcaccca ggccaccggc accgacgact gcaccggcga cgccgtggtg tacacgcacc 2040
gggcggggga gctcgtgcgg ctgccccggg gcgccaccct gccggtgacg ctcaagaggc 2100
tcgaatacga gctgttccac gtgtgccccg tccgcgccgt ggcgccggac atctcgttcg 2160
cgcccatcgg gttgctccac atgttcaatg ccggcggtgc cgtcgaggag tgcgtcgtca 2220
ggacgaacga ggacgacaag gccgttgtgg cgctcagggt gcgcggctgc ggccggttcg 2280
gcgcctactg ctcgcggagg ccggcgaaat gctccctcga ctcggctgac gtggagttcg 2340
gctacgacgc cgacacgggg ctcgtcacgg tcgacgtgcc ggtcccggag gaggagatgt 2400
accggtggac gctggagatt cgggtctagg ctggcatgct cgtgttcacg atcacggggc 2460
tggctgaggc gatctcagcc gttggattat ttctttccat tgttggtgcc tatagtgctt 2520
aagagagatt gatcgagaga taaatattgc catgtagaga gattatgctt tgctctatca 2580
ttgtggcaaa tgcaataatg tatcccttct gtactatcat taagatttga atgaatgcaa 2640
taaaggatgg tgctccttgt aaaaaaaa                                    2668
```

<210> 14
<211> 751
<212> PRT
<213> Triticum aestivum

<400> 14

```
Met Thr Val Thr Pro Gln Ile Thr Val Ser Asp Gly Arg Leu Ala Val
  1               5                  10                  15

Arg Gly Arg Thr Val Leu Ser Gly Val Pro Asp Asn Val Thr Ala Ala
             20                  25                  30

His Ala Ser Gly Ala Gly Leu Val Asp Gly Ala Phe Val Gly Ala Thr
         35                  40                  45

Ala Gly Glu Ala Lys Ser His His Val Phe Thr Phe Gly Thr Leu Arg
     50                  55                  60

Asp Cys Arg Phe Met Cys Leu Phe Arg Phe Lys Leu Trp Trp Met Thr
 65                  70                  75                  80

Gln Arg Met Gly Thr Ser Gly Arg Asp Val Pro Leu Glu Thr Gln Phe
                 85                  90                  95

Ile Leu Ile Glu Val Pro Ala Ala Ala Gly Asn Asp Asp Gly Asp Ser
            100                 105                 110

Glu Pro Val Tyr Leu Val Met Leu Pro Leu Leu Glu Gly Gln Phe Arg
        115                 120                 125

Thr Val Leu Gln Gly Asn Asp Gln Asp Gln Leu Gln Ile Cys Ile Glu
    130                 135                 140

Ser Gly Asp Lys Ala Val Gln Thr Glu Gln Gly Met Asn Ser Val Tyr
145                 150                 155                 160

Ile His Ala Gly Thr Asn Pro Phe Asp Thr Ile Thr Gln Ala Val Lys
                165                 170                 175

Ala Val Glu Lys His Met Gln Thr Phe His His Arg Glu Lys Lys Lys
            180                 185                 190

Val Pro Ser Phe Val Asp Trp Phe Gly Trp Cys Thr Trp Asp Ala Phe
        195                 200                 205
```

Tyr Thr Asp Val Thr Ala Asp Gly Val Lys Gln Gly Leu Arg Ser Leu
210 215 220

Ala Glu Gly Gly Ala Pro Pro Arg Phe Leu Ile Ile Asp Asp Gly Trp
225 230 235 240

Gln Gln Ile Gly Ser Glu Asn Lys Glu Asp Pro Ser Val Ala Val Gln
245 250 255

Glu Gly Ala Gln Phe Ala Ser Arg Leu Thr Gly Ile Lys Glu Asn Thr
260 265 270

Lys Phe Gln Ser Glu Gln Gln Glu Glu Thr Pro Gly Leu Lys Arg Leu
275 280 285

Val Glu Glu Thr Lys Lys Glu His Gly Val Lys Ser Val Tyr Val Trp
290 295 300

His Ala Met Ala Gly Tyr Trp Gly Gly Val Lys Pro Ser Ala Ala Gly
305 310 315 320

Met Glu His Tyr Glu Ser Ala Leu Ala Tyr Pro Val Gln Ser Pro Gly
325 330 335

Val Thr Gly Asn Gln Pro Asp Ile Val Met Asp Ser Leu Ser Val Leu
340 345 350

Gly Leu Gly Leu Val His Pro Arg Lys Val Tyr Ser Phe Tyr Asp Glu
355 360 365

Leu His Ala Tyr Leu Ala Ala Cys Gly Val Asp Gly Val Lys Val Asp
370 375 380

Val Gln Asn Ile Val Glu Thr Leu Gly Ala Gly His Gly Gly Arg Val
385 390 395 400

Ala Leu Thr Arg Ala Tyr His Arg Ala Leu Glu Ala Ser Val Ala Arg
405 410 415

Asn Phe Pro Asp Asn Gly Cys Ile Ser Cys Met Cys His Asn Thr Asp
420 425 430

Met Leu Tyr Ser Ala Lys Gln Thr Ala Val Val Arg Ala Ser Asp Asp
435 440 445

Phe Tyr Pro Arg Asp Pro Ala Ser His Thr Val His Ile Ser Ser Val
450 455 460

Ala Tyr Asn Thr Leu Phe Leu Gly Glu Phe Met Gln Pro Asp Trp Asp
465 470 475 480

Met Phe His Ser Leu His Pro Ala Ala Glu Tyr His Gly Ala Ala Arg
485 490 495

Ala Ile Gly Gly Cys Pro Ile Tyr Val Ser Asp Lys Pro Gly Asn His
500 505 510

Asn Phe Asp Leu Leu Lys Lys Leu Val Leu Pro Asp Gly Ser Val Leu
515 520 525

```
Arg Ala Gln Leu Pro Gly Arg Pro Thr Arg Asp Cys Leu Phe Ser Asp
    530                 535                 540

Pro Ala Arg Asp Gly Ala Ser Leu Leu Lys Ile Trp Asn Met Asn Lys
545                 550                 555                 560

Cys Ala Gly Val Val Gly Val Phe Asn Cys Gln Gly Ala Gly Trp Cys
                565                 570                 575

Arg Val Val Lys Lys Thr Arg Ile His Asp Glu Ala Pro Gly Thr Leu
            580                 585                 590

Thr Gly Ser Val Arg Ala Glu Asp Val Glu Gly Ile Thr Gln Ala Thr
        595                 600                 605

Gly Thr Asp Asp Cys Thr Gly Asp Ala Val Val Tyr Thr His Arg Ala
    610                 615                 620

Gly Glu Leu Val Arg Leu Pro Arg Gly Ala Thr Leu Pro Val Thr Leu
625                 630                 635                 640

Lys Arg Leu Glu Tyr Glu Leu Phe His Val Cys Pro Val Arg Ala Val
                645                 650                 655

Ala Pro Asp Ile Ser Phe Ala Pro Ile Gly Leu Leu His Met Phe Asn
            660                 665                 670

Ala Gly Gly Ala Val Glu Glu Cys Val Val Arg Thr Asn Glu Asp Asp
        675                 680                 685

Lys Ala Val Val Ala Leu Arg Val Arg Gly Cys Gly Arg Phe Gly Ala
    690                 695                 700

Tyr Cys Ser Arg Arg Pro Ala Lys Cys Ser Leu Asp Ser Ala Asp Val
705                 710                 715                 720

Glu Phe Gly Tyr Asp Ala Asp Thr Gly Leu Val Thr Val Asp Val Pro
                725                 730                 735

Val Pro Glu Glu Glu Met Tyr Arg Trp Thr Leu Glu Ile Arg Val
            740                 745                 750
```

<210> 15
<211> 2653
<212> DNA
<213> Triticum aestivum

<400> 15

```
tggcgcgcac agcactcatc tccagctgcc gcagcaccgg ggccgccgca tcctcctcgt   60
ctccctctcc tcgcgcgcgt atatataccc gcgcctcctc ttcctcttcc cgcccaccgc  120
tcctcgcctc tccgcctcct ccccgatcca ctcctcgtcg gtgttcgatt aagactcttg  180
cggcggtcaa gggagcgtcg ttgtccggct ggcggagcgc gagggaggag aaggagttgg  240
agatgacgat cgagtcgtcc gtgcggctcg ccggcgggga gctgtcggtc cgcgggcgca  300
cggtgctgtc cggcgtgccg gacgcggtgt cggcgtcgcc cgcggcggcc cggggccccg  360
tcgacggcgt cttcctcggc gccgacctcg ccggccccgc ctcccgccac gtcgtctccc  420
tcggccacat gaggggcgtg cggttcatgg cgtgcttccg gttcaagatg tggtggatgg  480
cgcagaggat gggggacaag ggcggggacg tcccgcacga gacgcagttc ctgctggtgg  540
agtccagggc catcggcggc gaggaggacg acgcgtcgta cgtcgtgttc ctcccgctcg  600
tggagggcgc gttccgggcc agcctccagg gcggcggcgc gggcggcgac gagctccagc  660
```

tctgcgtcga gagcggcgac gccggcacgc tcgccagttc cttcgaccgc gcgctcttcg 720
tgggcgccgc ggactccgac ccgttcgccg ccatcgccgg cgccgtcgcg gccgtcaggt 780
cctgcctcgg gaccttccgc ccgcgcgccg agaagaagct ccccgccatc gttgactact 840
tcgggtggtg cacgtgggac gccttctacc aggacgtcac gcaggagggc gtcgaggccg 900
ggctccagag cctcgccgcc ggcggagctc cgcccaagtt cgtcatcatc gacgacggct 960
ggcagtcggt cgggaccgac aaacagagcc ccgacttgga ctctgcgggc gaggccggca 1020
agtcgccgcc ccttccccgg ctcaccggca tcaaggagaa cagcaagttc cagagcggcg 1080
acgacccggc caccgccacc ggcatcgaga cgctggtgcg cgcggccaag gagaagtacg 1140
ggctcaagta cgtgtacgtc tggcacgcca tcaccggcta ctggggcggc gtgcggccgg 1200
gcgtcgccgg gatggaggcg tatcgctcgt ccatgcagtt ccccaagatc tcgccgggcg 1260
tggcggagaa cgagcccaac atgaagaccg acgtgctcac cctgcagggg ctcggcctcg 1320
tgcacccgca ggccgtgcac cgcttctacg acgagctcca cgcgtacctc gccgccgccg 1380
gcgtcgacgg cgtcaaggtg gacgtgcagt gcgtcctcga gacgctcggc gccggccacg 1440
gcggccgcgt gcagctcacc agggagtacc accgcgcgct cgacgcctcc gtcgccaaga 1500
acttccccga caacggcatc atcgcctgca tgagccacaa caccgacgcc ctctactgct 1560
cgaagcagac ggcggtggtg agagcgtcgg acgatttctt cccgagggag gcggtgtcgc 1620
acacgatcca catcgcggcg gtggcgtaca acagcgtgtt cctcggcgag ttcatgctcc 1680
cggactggga catgttccac tccctccacc ccgccggcga ctaccacggc tcggcgcgcg 1740
ccatcagcgg cgggcccgtg tacgtcagcg acgcgccggg gaagcacgac ttcgagctgc 1800
tgaggaagat ggtgctgccg gacggcaccg tgctgcgcgc gcggctgccg ggccggccga 1860
ccagggactg cctgttcgcg gacccggcgc gcgacggcgc caccctgctc aagatctgga 1920
acatgaacag gttcacgggc gtgctcggcg tgtacaactg ccagggcgcg gcgtggagct 1980
ccgcggagaa gaagaacgtg ttccaccagg aggccggcgc cggcgccctg acctgcggcg 2040
tcaggggccg cgacgtccac ctcatcgccg aggccgcgac ggacggcggc gccgggtgga 2100
gcggcgactg cgccgtgtac cgccacggcg ccggcgacct cgtggtgctc cccgacggcg 2160
tggcgctgcc cgtgtccctc aaggtcctgg agcacgacgt gctcaccgtg tcgccgatca 2220
aggatttggc ggccgggttc aggttcgcgc cggtcggcct cgtggacatg ttcaacggcg 2280
gcgcggcggt ggaaggcctg acctacagcc tccttgccga cggcgaggag gcggtcgggc 2340
tggtgagcat ggaagtgcga gggcgcggga ggttcggcgc ctactcgtcg gtccggccga 2400
ggagttgcac gctggggtca gccccggcgg agttctccta cgacgcctcc tccggcatgg 2460
tgatcctcga gctcgagtcc atgccattgc ccaaggaaag ggttcacaag atcgccattg 2520
agctgtagaa tgttttagct ccccggtcat gcgatccatc gatagtgtat gctttcgtga 2580
ttattattga attgaagggt tggggaaata aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2640
aaaaaaaaaa aaa 2653

<210> 16
<211> 841
<212> PRT
<213> Triticum aestivum

<400> 16
Ala Arg Thr Ala Leu Ile Ser Ser Cys Arg Ser Thr Gly Ala Ala Ala
1 5 10 15

Ser Ser Ser Ser Pro Ser Pro Arg Ala Arg Ile Tyr Thr Arg Ala Ser
20 25 30

Ser Ser Ser Ser Arg Pro Pro Leu Leu Ala Ser Pro Pro Pro Pro Arg
35 40 45

Ser Thr Pro Arg Arg Cys Ser Ile Lys Thr Leu Ala Ala Val Lys Gly
50 55 60

Ala Ser Leu Ser Gly Trp Arg Ser Ala Arg Glu Glu Lys Glu Leu Glu
65 70 75 80

Met Thr Ile Glu Ser Ser Val Arg Leu Ala Gly Gly Glu Leu Ser Val
85 90 95

```
Arg Gly Arg Thr Val Leu Ser Gly Val Pro Asp Ala Val Ser Ala Ser
            100                 105                 110

Pro Ala Ala Ala Arg Gly Pro Val Asp Gly Val Phe Leu Gly Ala Asp
        115                 120                 125

Leu Ala Gly Pro Ala Ser Arg His Val Val Ser Leu Gly His Met Arg
    130                 135                 140

Gly Val Arg Phe Met Ala Cys Phe Arg Phe Lys Met Trp Trp Met Ala
145                 150                 155                 160

Gln Arg Met Gly Asp Lys Gly Gly Asp Val Pro His Glu Thr Gln Phe
                165                 170                 175

Leu Leu Val Glu Ser Arg Ala Ile Gly Gly Glu Glu Asp Asp Ala Ser
            180                 185                 190

Tyr Val Val Phe Leu Pro Leu Val Glu Gly Ala Phe Arg Ala Ser Leu
        195                 200                 205

Gln Gly Gly Gly Ala Gly Gly Asp Glu Leu Gln Leu Cys Val Glu Ser
    210                 215                 220

Gly Asp Ala Gly Thr Leu Ala Ser Ser Phe Asp Arg Ala Leu Phe Val
225                 230                 235                 240

Gly Ala Ala Asp Ser Asp Pro Phe Ala Ala Ile Ala Gly Ala Val Ala
                245                 250                 255

Ala Val Arg Ser Cys Leu Gly Thr Phe Arg Pro Arg Ala Glu Lys Lys
            260                 265                 270

Leu Pro Ala Ile Val Asp Tyr Phe Gly Trp Cys Thr Trp Asp Ala Phe
        275                 280                 285

Tyr Gln Asp Val Thr Gln Glu Gly Val Glu Ala Gly Leu Gln Ser Leu
    290                 295                 300

Ala Ala Gly Gly Ala Pro Pro Lys Phe Val Ile Ile Asp Asp Gly Trp
305                 310                 315                 320

Gln Ser Val Gly Thr Asp Lys Gln Ser Pro Asp Leu Asp Ser Ala Gly
                325                 330                 335

Glu Ala Gly Lys Ser Pro Pro Leu Pro Arg Leu Thr Gly Ile Lys Glu
            340                 345                 350

Asn Ser Lys Phe Gln Ser Gly Asp Asp Pro Ala Thr Ala Thr Gly Ile
        355                 360                 365

Glu Thr Leu Val Arg Ala Ala Lys Glu Lys Tyr Gly Leu Lys Tyr Val
    370                 375                 380

Tyr Val Trp His Ala Ile Thr Gly Tyr Trp Gly Gly Val Arg Pro Gly
385                 390                 395                 400

Val Ala Gly Met Glu Ala Tyr Arg Ser Ser Met Gln Phe Pro Lys Ile
                405                 410                 415
```

```
Ser Pro Gly Val Ala Glu Asn Glu Pro Asn Met Lys Thr Asp Val Leu
            420                 425                 430

Thr Leu Gln Gly Leu Gly Leu Val His Pro Gln Ala Val His Arg Phe
        435                 440                 445

Tyr Asp Glu Leu His Ala Tyr Leu Ala Ala Ala Gly Val Asp Gly Val
    450                 455                 460

Lys Val Asp Val Gln Cys Val Leu Glu Thr Leu Gly Ala Gly His Gly
465                 470                 475                 480

Gly Arg Val Gln Leu Thr Arg Glu Tyr His Arg Ala Leu Asp Ala Ser
                485                 490                 495

Val Ala Lys Asn Phe Pro Asp Asn Gly Ile Ile Ala Cys Met Ser His
            500                 505                 510

Asn Thr Asp Ala Leu Tyr Cys Ser Lys Gln Thr Ala Val Val Arg Ala
        515                 520                 525

Ser Asp Asp Phe Phe Pro Arg Glu Ala Val Ser His Thr Ile His Ile
    530                 535                 540

Ala Ala Val Ala Tyr Asn Ser Val Phe Leu Gly Glu Phe Met Leu Pro
545                 550                 555                 560

Asp Trp Asp Met Phe His Ser Leu His Pro Ala Gly Asp Tyr His Gly
                565                 570                 575

Ser Ala Arg Ala Ile Ser Gly Gly Pro Val Tyr Val Ser Asp Ala Pro
            580                 585                 590

Gly Lys His Asp Phe Glu Leu Leu Arg Lys Met Val Leu Pro Asp Gly
        595                 600                 605

Thr Val Leu Arg Ala Arg Leu Pro Gly Arg Pro Thr Arg Asp Cys Leu
    610                 615                 620

Phe Ala Asp Pro Ala Arg Asp Gly Ala Thr Leu Leu Lys Ile Trp Asn
625                 630                 635                 640

Met Asn Arg Phe Thr Gly Val Leu Gly Val Tyr Asn Cys Gln Gly Ala
                645                 650                 655

Ala Trp Ser Ser Ala Glu Lys Lys Asn Val Phe His Gln Glu Ala Gly
            660                 665                 670

Ala Gly Ala Leu Thr Cys Gly Val Arg Gly Arg Asp Val His Leu Ile
        675                 680                 685

Ala Glu Ala Ala Thr Asp Gly Gly Ala Gly Trp Ser Gly Asp Cys Ala
    690                 695                 700

Val Tyr Arg His Gly Ala Gly Asp Leu Val Val Leu Pro Asp Gly Val
705                 710                 715                 720

Ala Leu Pro Val Ser Leu Lys Val Leu Glu His Asp Val Leu Thr Val
                725                 730                 735
```

```
Ser Pro Ile Lys Asp Leu Ala Ala Gly Phe Arg Phe Ala Pro Val Gly
            740                 745                 750

Leu Val Asp Met Phe Asn Gly Gly Ala Ala Val Glu Gly Leu Thr Tyr
        755                 760                 765

Ser Leu Leu Ala Asp Gly Glu Glu Ala Val Gly Leu Val Ser Met Glu
    770                 775                 780

Val Arg Gly Arg Gly Arg Phe Gly Ala Tyr Ser Ser Val Arg Pro Arg
785                 790                 795                 800

Ser Cys Thr Leu Gly Ser Ala Pro Ala Glu Phe Ser Tyr Asp Ala Ser
                805                 810                 815

Ser Gly Met Val Ile Leu Glu Leu Glu Ser Met Pro Leu Pro Lys Glu
            820                 825                 830

Arg Val His Lys Ile Ala Ile Glu Leu
        835                 840
```